

Help! The data are coming

Some branches of science have learnt how to cope with huge amounts of information. Biologists haven't. There is a dearth of essential skills which is only now starting to be taken seriously.

Terabytes, exabytes, yottabytes: bytes by the zillion are heading this way, thanks to the appetites of particle physicists, Earth scientists and organismal and molecular biologists (see Briefing on pages 517–520). So how will researchers cope?

The particle physicists are well versed in handling such data. They seem to be crossing their fingers and hoping that the off-the-shelf hardware and software will be available in time to allow them to analyse the copious output of their next-generation colliders. Given the history of the technology, their expectations are well founded.

Molecular biologists, on the other hand, appear to have eyes for data that are bigger than their stomachs. As genomes near completion, as DNA arrays on chips begin to reveal patterns of gene sequences and expression, as researchers embark on characterizing all known proteins, the anticipated flood of data vastly exceeds in scale anything biologists have been used to.

Biologists are waking up to the challenge, albeit belatedly. Last week, for example, an expert panel told Harold Varmus, director of the US National Institutes of Health (NIH), that his agency needs to do much more in this direction. As panel co-chair David Botstein rightly emphasized, the big issue is training. Nevertheless, sheer supercomputer power (and, for that matter, medium-sized computer power, too) will be essential. To that end, the panel recommended support for the development of existing supercomputing facilities established for other disciplines, as opposed to new facilities for biology. The strain on existing facilities is enormous, so serious investment will be required, but, the panel argues, not in a way that

re-invents wheels. This approach also has the advantage of leveraging NIH dollars at centres that already house needed expertise and infrastructure.

But, equally urgently, the NIH needs to help develop a new generation of computer-wise researchers. The panel recommended that five to twenty “National Programs of Excellence in Biomedical Computing” be established. It’s a measure of the speed of the revolution, and of the dearth of expertise in the community, that it’s not obvious where in the United States such programmes would be launched. Just as urgent is the need for computer specialists in laboratories, and a change in attitude that sees them as invaluable rather than second-class citizens. In short, computer experts at \$85,000 a year are, like it or not, an increasingly necessary component of grant applications.

It would be wrong to leave the US agenda in this area solely in the hands of the NIH. The National Science Foundation has traditionally been the focus of support for supercomputing in the research community, while the Department of Energy has supported most of the country’s particle physicists. Both agencies have made their own investment in informatics, and sharing what they learn with biologists is as urgent a necessity as increasing NIH spending.

Other countries and regions face similar shortages of skills. Meanwhile, private companies are busily sequencing and computing and licensing. There is, therefore, an urgent underlying message that all scientifically ambitious countries should heed: strong government funding for the quantitative analysis of data is essential if the results of fundamental biological research are to remain a public good. □

A cause worth funding

A German synchrotron would be good for the Middle East.

It’s too easy for *Nature* to urge the world to spend more money on science. On the whole, that temptation is resisted. But there are honourable exceptions. A proposal — as yet unfunded — to establish a joint synchrotron radiation facility in the Middle East is one such, and deserves immediate attention.

The government of Germany is understood to be receptive to the idea of giving away a fully functioning synchrotron radiation source for use by scientists in the Middle East (see pages 507–508). The synchrotron is to be the focus of a broader centre for research excellence for scientists from throughout the region, as well as other parts of the world. The project’s founders envisage a facility similar in aim to the European Laboratory for Particle Physics (CERN), which brought together scientists from countries that had fought each other during the Second World War.

Scientists nominated by many of the region’s governments will discuss the project at a meeting organized by the United Nations Educational, Scientific and Cultural Organization in Paris next week. Israel is expected generously to agree not to bid to host the synchrotron — as its scientific competence would well qualify it to — allowing the facility to be housed in one of its neighbouring countries. There appears to be no shortage of potential hosts, with Cyprus, Egypt and the Palestinian Authority among the contenders.

But the proposal needs funds in no small measure. There are several potential sources. These include the European Union and the US government, as well as states within the Middle East itself. The issue of funds for the project will also be raised at the World Conference on Science in Budapest later this month. *Nature*’s advice to any potential funder is not to hold back, for this will be a worthwhile investment. Initiatives such as this do not come around often. When they do, they should be supported unhesitatingly.

After a troubled half-century, the peoples of the Middle East are making the slow transition to peace. It is sometimes hard to imagine, but there was a time not so long ago when the Christians, Jews and Muslims of the Middle East lived in relative harmony, when philosophers and scientists were recruited to the region’s leading institutions of learning because of their expertise, and not on the basis of their faith or geographic identity.

Is it too optimistic to suggest that next week’s meeting in Paris may mark the return of such happier times? Probably. But the meeting will be a valuable and long-awaited beginning. And if the project succeeds, it could be a step closer to the day scientists from Israel and its neighbours are free to travel to — and work in — one another’s laboratories, exchange information and cooperate in research. That alone would be a major step forward. □



Middle East synchrotron facility could bring regional cooperation

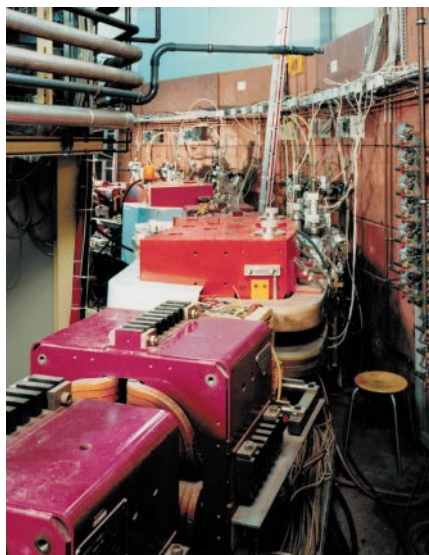
[LONDON] Scientists and science administrators from the Middle East, Europe, the United States and the Far East are to meet in Paris next week to discuss plans to set up a joint synchrotron radiation facility, which would be the first regional centre for cooperation in basic research in the Middle East.

The meeting will be attended by scientists from the region, including Cyprus, Egypt, Iran, Israel, Jordan, Lebanon, the Palestinian Authority, Syria and Turkey. Two officials from the US administration are also expected to attend, along with officials from the Abdus Salam Centre for Theoretical Physics in Trieste, Italy.

Plans for the new centre revolve around the strong probability that Germany is prepared to present scientists in the Middle East with BESSY-1, a 0.8 GeV synchrotron radiation source as a gift. BESSY-1, currently based in Berlin, is still active after 14 years, but is due to be shut down at the end of the year. No country in the Middle East has such a facility.

The synchrotron, which is worth \$60 million, will be replaced by BESSY-2. An upgraded 1 GeV BESSY-1 would form the core of a planned international centre of excellence providing training and research for the Middle East's scientists in structural biology, environmental and materials science.

A key aim will also be to attract scientists from Europe and the United States to the region. The facility may in addition tempt back some of the Middle East's many scientists



Young at heart: after 14 years in Berlin, BESSY (above) may be heading for a second lease of life in the Middle East.

who have left to work abroad. The facility will also have an important diplomatic role: helping to improve relations within one of the world's more troubled regions.

The project's founders are Herman Winick of the Stanford Linear Accelerator Centre in California, a member of the Machine Advisory Committee of BESSY-2, and fellow committee member Gustav-Adolf Voss, a former director at the DESY laboratory in Hamburg.

Winick is credited with coming up with the idea of moving BESSY-1 to the Middle East during discussions about the future of the machine. The idea was taken further through the Middle East Scientific Collaboration network (MESC), a network of scientists promoting research cooperation between Europe, the United States and the Middle East.

Participants at next week's meeting in Paris will discuss the facility's governing apparatus and a timetable for action. Herwig Schopper, former director-general of the European Laboratory for Particle Physics (CERN) and an active member of MESC, will chair the meeting.

The United Nations Educational, Scientific and Cultural Organization (Unesco) is expected to be confirmed as the facility's umbrella body, mainly because of its long experience and perceived neutrality in the region. Participants at the meeting will also be urged to lobby their governments to support the facility at the World Conference on Science in Budapest later this month. They will aim to have the project included in the conference's Framework for Action document.

Funding for the project will be high on the agenda. Despite not having to pay for BESSY-1 itself, around \$10 million will be needed for the upgrade, \$10 million for construction and infrastructure, and about \$10 million a year will go on running costs.

There are no obvious sources for the money. Cash-constrained Unesco is expected to provide little more than important diplomatic support, and the Middle East's richer countries, such as the oil-producing states of Saudi Arabia and Kuwait, have so far chosen not to participate.

The project's founders are not expecting Germany to fill the shortfall, partly out of a feeling that it has already done enough. But Germany will be lobbied to help obtain European Union funds before its presidency of the body ends this month.

The influential US physical-science community will be urged to investigate potential sources in the United States. One such source could be the estimated \$2 billion Middle East aid package currently being negotiated in the US Congress. In addition, the US Department of Energy is understood to be receptive to the idea of training scientists from the Middle East in accelerator technology and the uses of synchrotron radiation sources at its labs.

But the Paris meeting will not take a decision on the most sensitive issue: the synchrotron's location. Instead, governments who want to host the facility will be

Pakistan prime minister pledges science boost

[LONDON] Pakistan's prime minister, Mian Nawaz Sharif, last month announced plans for a 1 billion rupee (US\$23 million) investment in science and technology. The announcement was made on the first anniversary of Pakistan's nuclear tests.

About half the money is expected to be spent on research and research infrastructure, including scholarships for 300 additional PhD students in Pakistani universities, and for 100 students to be sent overseas. The remaining funds will be spent on science education.

Sharif has pledged to

provide a science teacher for every government school, and has promised to set up polytechnics in each of the country's administrative divisions. The government will also set up a committee to recommend ways of giving more autonomy to research institutions.

"One year ago we broke the shackles of dependence on foreign technology and conducted successful nuclear tests," said Sharif. "We should now focus our attention on making Pakistan an economic power."

The initiatives are the most comprehensive and ambitious that Pakistan has

seen for a decade. They have also come as a surprise for many in Pakistan's science community because the present Muslim League government – like its predecessor – had previously shown little interest in science.

Many scientists are uncertain whether the funds will materialize, as similar promises were not honoured in the past. But one senior scientist says he believes this initiative is different, as it originates from the prime minister's circle of close advisers, rather than from the science ministry.

E.M.

invited to submit bids to its governing body. Cyprus, Egypt and the Palestinian Authority have indicated that they will do so, and their bids are being taken seriously. Other potential hosts, such as Turkey, may emerge after the meeting.

Despite being the Middle East's most developed country scientifically, Israel appears to have decided not to bid. This is largely because an Israel-based facility would end any hope of participation from scientists from most of the region's other countries.

"The advantages of having it in an Arab country are greater than having it in Israel," says Eliezer Rabinovici, professor of physics at the Hebrew University of Jerusalem. Scientists from Iran, Lebanon and Syria would not have been allowed by their governments to participate if there had been a chance that the facility would find a home in Israel.

But Israel's active participation is considered to be important for the project's success, says Voss.

The Palestinian Authority is a strong contender to host the synchrotron. The project's founders, together with Ernst Wehreter, a senior scientist at BESSY, have made visits to Cyprus, Israel, Jordan and the West Bank. The West Bank's chief advantages include its accessibility to scientists from both Israel and the rest of the Middle East, and a centre of scientific excellence would give a much-needed boost to Palestinian scientists, if not the Israeli-Palestinian peace process.

"Our scientific growth has been stunted for many years, and we need opportunities to help build our infrastructure so that we can join the modern world," says Hanna Hallak, dean of science and chairman of the physics department at Bethlehem University.

But a Palestinian bid has two prerequisites, which may not be easy to obtain. The first is continued peace between the government of Israel and its Palestinian population — renewed unrest will make it difficult for scientists from other countries to work at the centre. A second prerequisite is that most of the funding will need to be found overseas, as the Palestinian Authority is unlikely to contribute substantially to the facility.

This is where the supporters of Egypt's bid believe that they could have an edge. Egypt remains more stable politically than its neighbour. The government is also prepared to invest considerable sums towards the facility's construction and running costs, says Hamid Roushdy El-Kady, emeritus professor of radiation biology at the National Centre for Radiation Research and former chairman of Egypt's Atomic Energy Authority.

Egypt has been developing accelerator technology since 1961, says El-Kady, adding that most of the region's non-Israeli scientists go to Egypt for training in radiation biology and high-energy physics.

Ehsan Masood

Biotech panel set up in US may help allay public fears

[WASHINGTON] The US Department of Agriculture (USDA) is to set up a biotechnology advisory committee made up of scientists, farmers, industry groups, environmentalists and members of the public. The move has been made partly in the hope that it may head off the kind of controversy over genetically modified (GM) foods that has raged recently in Europe.

The department plans to begin soliciting nominations for members this week. The agriculture secretary, Dan Glickman, who announced the new committee in March, will select 25 people representing a range of viewpoints. According to Michael Schechtman of the USDA, who will be the panel's executive secretary, the committee will be expected to meet a maximum of four times a year.

Environmentalists have praised USDA's new willingness to listen to public concern about agricultural biotechnology. While cautioning that the committee's goals are "still a little unclear", Rebecca Goldburg of the Environmental Defense Fund in New York says that it represents "a real change".

"I think Glickman has been gaining an understanding that he and industry cannot talk this problem away," says Jane Rissler of the Union of Concerned Scientists. USDA, along with the Environmental Protection Agency (EPA) and the Food and Drug Administration, has regulatory responsibility for GM foods.

Glickman appears to have backed away recently from what some have characterized as unabashed 'boosterism' about GM crops. "We can't force these new genetically engineered food products down consumers' throats," he said in April at a commencement ceremony at Purdue University, Indiana. "Dismissing the scepticism that's out there is not only arrogant, it's also a bad business strategy," he added.



Glickman: 'let's win over our opponents'.

Regarding the decision by some British supermarket chains to ban GM foods, the secretary said that such organizations needed "a little bit of educating, but I don't think we can just sit here and berate them".

He struck a similar tone before the World Agricultural Congress in St Louis in May, saying: "Americans are more willing to see science as a force for progress... while Europeans may be more cautious, more concerned perhaps about even the theoretical possibility of risk."

But the USDA panel's membership may be contentious, if a recent fight over the composition of a National Academy of Sciences committee on pesticide-resistant GM plants is an indication of how polarized the debate has become (see *Nature* 399, 7; 1999).

Attempts to bring all the 'stakeholders' around the same table can backfire badly, as in the case of a 50-member panel formed by the EPA, at the request of the vice-president, Al Gore, to discuss pesticides in food. In April, seven environmental groups walked out of the Tolerance Reassessment Advisory Committee, claiming that it was dominated by agricultural and chemical interests.

The EPA is holding public meetings this summer to discuss pest-resistance management plans for GM crops.

But there has been little progress on a proposal to raise GM crops as a trade issue at a meeting of G8 nations this summer, which some in Congress have been pushing (*Nature* 399, 287; 1999).

Tony Reichhardt

German plan to curb rise in research funds

[MUNICH] Germany's 16 *Länder* (states) say they want to cut down considerably on the five per cent increase in next year's budget that has been promised to the main basic research organizations, the Deutsche Forschungsgemeinschaft (DFG) and the Max Planck Society (MPS).

The DFG and MPS are jointly funded by the federal government and the *Länder*. A recent report from the German Press Agency quotes an unpublished agreement drawn up by the *Länder* ministers of finance to limit the budget increase to two per cent in 2000. The agreement also says that budgets should

be negotiated annually in future, rather than guaranteeing fixed mid-term increases.

If approved by the *Länder* this month, the plan could hinder implementation of the recommendations of an external committee that evaluated internal procedures of the MPS and DFG (see *Nature* 399, 395–396; 1999). The committee said that a guaranteed mid-term financial framework should be maintained to allow reforms to be made. But the federal government is unlikely to block a cost-cutting *Länder* agreement at a time when the federal budgets for 2000 are to be reduced by 7.4 per cent.

Quirin Schiermeier

AP/BRIAN HENDRICKSON

NASA tells physicists to aim for the stars

[WASHINGTON] Dan Goldin, the administrator of the US space agency NASA, has triggered fierce debate among physicists by suggesting that the next generation of instruments for exploring the fundamental nature of matter should be built in space.

Speaking to a joint meeting of astrophysicists and particle physicists at Chicago-based Fermilab, the largest US particle-physics laboratory, Goldin said that "space will be an essential laboratory in our journey beyond the Standard Model".

He told the meeting that Enrico Fermi himself had once proposed an "accelerator in space". "Fermi had it right," he continued. "Space is the way to go for the future. Fermi was a dreamer, and that's what we should be too. Fermi hadn't seen the physical and economic limitations of ground-based accelerators, but he made the leap to space anyway."

Goldin's remarks caused some consternation at Fermilab, which last week dedicated a new main injector to the world's most powerful hadron accelerator, the Tevatron, and which now needs to develop a case for a new accelerator based firmly on the ground in Illinois (see *Nature* 394, 611; 1998).

This discomfort was compounded, according to laboratory officials, by additional comments by Goldin, intimating that NASA's experience in searching for life in the Universe could assist high-energy physics, "because it's dead". (NASA officials said later that Goldin was referring only to the high-energy physics programme at NASA itself.)

However, Goldin's primary objective was not to embarrass his hosts but to tell the scientific community that NASA is prepared to spend big money on physics in space. The agency also wants to work more closely with the Department of Energy and the National Science Foundation (NSF) to support space-based physics experiments.

NASA officials say that the agency envisages that a new generation of such instruments will give the agency a broader mission in astrophysics and cosmology. Money for these projects could start to flow as soon as the 2002 fiscal year, the officials say.

Goldin plans to sell these experiments as part of a 'cosmic origins' programme. Introducing the concept in Chicago, he compared it to NASA's Origins programme to determine the origins of life, which he launched three years ago and which, he pointed out, is now to be funded at \$1 billion over five years.

Although Goldin's speech avoided direct criticism of ground-based approaches to particle physics, he did describe space-based observation of gravity waves, cosmic rays and new high-energy particles as "the gateway to the new physics".

Goldin illustrated the approach by citing instruments that already exist as conceptual



programmes at NASA, such as the Laser Interferometer Space Antenna, which would detect gravity waves by measuring tiny changes in the distance between satellites positioned five million kilometres apart.

The terrestrial Laser Interferometer Gravitational-wave Observatory, currently being built by the NSF to do this, has arms that are 4 km apart, and its accuracy is limited by seismic noise.

Scientists who advised Goldin on the speech say that it was meant to provoke researchers to think up new ideas for space-based physics instruments. "Goldin is challenging the community to find ways to do fundamental physics in space," says Rocky Kolb, a cosmologist at Fermilab.

Varmus relaunches NIH degree proposal

[WASHINGTON] Harold Varmus, director of the US National Institutes of Health (NIH), made a second attempt last week to persuade his advisory committee that the institutes should award degrees, and establish a PhD programme in "disease-oriented integrative biology".

The committee was first informed of the proposal, under which 15 students would be admitted each year, last December. In a straw vote last week, most members of the committee backed further development of the proposal.

But several remained unconvinced that the NIH should launch a PhD programme when, it has been argued, US universities are producing a glut of life scientists. Such a move would "send the wrong symbolic message," said Shirley Tilghman, a committee member who is a professor of molecular biology at Princeton University.

Tilghman chaired a panel last year for the National Research Council on the early careers of life scientists that recommended freezing the number of life sciences PhD students. It said that new PhD programmes should be started only in "rare and special

"He doesn't have any particular experiments in mind, but he senses an opportunity," says Kolb. "Scientists are conservative in that way — we have to be encouraged to dream." But he concedes that "scientists like to work on instruments that might be built one day".

Alan Bunner, who heads the section of NASA's office of space science that supports astrophysics and cosmology, says the agency is considering greater emphasis on the area as part of its strategic plan for space science.

Bunner adds that NASA would like to work more closely with the Department of Energy and the NSF to support experiments. "It is time for a new approach, bringing all three agencies together," he says.

The energy department and NSF are already jointly assessing astrophysics proposals through a panel called SAGENAP — the scientific assessment group for experiments in non-accelerator physics, chaired by Barry Barish of Caltech. It is not clear if NASA intends to join up and help to pay for experiments approved by the group.

Bob Eisenstein, head of the mathematical and physical sciences directorate at NSF, says Goldin "gave a very exciting and far-ranging speech" at Fermilab. Eisenstein adds that space has obvious advantages as a venue for many types of experiment: "Based on intellectual promise, it is the best way to go," he says. "But when you take cost into account, it is pretty damn expensive." **Colin Macilwain**

circumstances" (see *Nature* 395, 103; 1998).

Varmus disagreed last week with Tilghman, who argued that an NIH programme would be warranted only if it taught material not available anywhere else. "Things that are done at NIH do not have to be unique," Varmus said.

The PhD programme proposal was presented by Michael Gottesman, NIH deputy director for intramural research. The curriculum would focus on "special needs" in US graduate education, including bioinformatics and clinical research.

Enthusiasts for the proposal included Eric Kandel of the Center of Neurobiology and Behavior at Columbia University in New York. "Focusing on programmes that you can do better than anybody else is a terrific idea," he said. "There are very few places in the country that can do this anywhere as well as NIH," he said of teaching bioinformatics and clinical investigation.

According to Varmus, work will continue on the proposal, and the advisory committee will receive a further report in December. **Meredith Wadman**



Lawton: 'no conspiracy against marine science.'

Lawton named head of UK environment research agency

[LONDON] John Lawton, director of the Centre for Population Biology at Imperial College, London, has been appointed to head the Natural Environment Research Council (NERC), Britain's main government funding agency for environmental research.

Lawton, who is already a member of NERC's governing council, will take up his post, initially for three years, in October. He replaces Sir John Krebs, who has been NERC's chief executive for the past five years.

The appointment has in general been widely welcomed. Lawton commands considerable respect, both as an ecologist — he has had 17 papers published in *Nature* and four in *Science* — and for his role as founding director of the Centre for Population Biology, which was set up 11 years ago.

But some members of the marine-science community say they are disappointed that a marine scientist was not appointed to head the research council. Lawton will be the second consecutive chief executive to have a background in zoology. Marine scientists feel that someone sharing their background would have been better motivated to lobby the government to raise funds for marine research, which has not fared as well as other environmental disciplines in recent years.

Lawton says he is aware of these concerns but, while declining to comment on them directly, he strongly denies any charges of a conspiracy against marine sciences in NERC.

He also says he is not planning any drastic changes to NERC's published three-year research agenda. But he adds that he is looking forward to developing interdisciplinary ideas within some of the agreed priority areas. These include exploring links between the environment and health, and between ecology and economics.

Lawton has a completely different personality to his predecessor. Krebs is noted for his diplomatic skills, whereas Lawton is usually unafraid to voice strong opinions, and has been described by one of his peers as "very persistent".

Ehsan Masood

French research boosted by grants run by ministries

[PARIS] France last week awarded life sciences, information technology and social sciences the lion's share of a recently created FF1.2 billion (US\$188 million) special annual fund for research. The fund is administered directly by government ministries, not by the research agencies that usually distribute grants.

The money comes from a new fund for basic science, the National Science Fund, and the existing Fund for Technology Research, which had fallen into neglect before being refloated by the current government. Grants will be distributed through calls for proposals, reviewed by ministerial committees.

Six priorities were earmarked for extra support at a meeting of the interministerial committee for scientific and technological research, chaired by prime minister Lionel Jospin, and attended by 12 ministers whose responsibilities include aspects of research.

Life sciences won FF395 million, with FF80 million going to the National Genotyping Centre at Evry near Paris and FF60 million to the new national plant biotechnology programme, Génoplante. Information technology will receive FF175 million, and FF55 million will go to nanotechnology. Human and social sciences received an injection of FF85 million, urban studies FF75 million,

Earth and environmental sciences FF45 million, and energy research FF20 million.

A further FF100 million was earmarked for a competition to fund promising ideas for the creation of high-technology companies. And FF40 million will create a fund for young researchers, with the money distributed through a call for proposals for original ideas irrespective of discipline.

The increased prominence of central funding has come under fire from trade unions representing scientists, who are suspicious of grants being administered directly by government in a country where research has long been under the jurisdiction of independent research agencies.

But Claudine Laurent, an adviser to science minister Claude Allègre, says the science ministry has a duty to stimulate and coordinate "the priorities of the government", arguing that the research agencies still have considerable freedom, and that the government funding supports fundamental research.

Olivier Kahn, a chemist at the University of Bordeaux, welcomes what he says is a clear hierarchy of national research priorities, and an effort to redress well-known weaknesses of the French research system, such as insufficient attention to the needs of young scientists.

Eric Glover

Funds promised for South African telescope

[CAPE TOWN] Construction of the Southern African Large Telescope (SALT), which is planned to become the largest single telescope in the Southern Hemisphere, has moved closer with commitments to joint funding from overseas partners in Poland, the United States and Germany.

The South African government announced its commitment to provide half the funding for the telescope a year ago (see *Nature* 393, 403; 1998). With 84 per cent of the expected cost of \$16.4 million (without instrumentation) now committed, construction of the telescope will commence later this year. The telescope is expected to take five years to complete.

The largest overseas contributor so far is the Polish Ministry of Science, which has said it will provide \$2.5 million for the construction, with an extra \$0.5 million from a consortium of Polish universities and research institutions. A further \$1.2 million for construction costs and \$1 million for running costs over the first ten years has been promised by Rutgers University in the United States.

The third overseas partner is the

University of Göttingen in Germany, which has committed \$1.3 million towards the telescope's construction. Prospective partners that have not yet committed funds include Carnegie Mellon University, Iowa State University and the University of Wisconsin in the United States, and the astronomical community in New Zealand.

SALT will have an effective diameter of 9.2 metres. Like its twin in the Northern Hemisphere, the Hobby-Eberly Telescope, recently built in Texas, SALT will be able to collect and analyse the light from many celestial objects at a time over a field of view of approximately 5 arcminutes.

Conventional telescopes need to move to track objects in the sky as the Earth rotates, but SALT will remain stationary, with a 5.5-tonne 'tracker assembly' at the top of the telescope moving instead. This design limits observations to the time it takes a celestial object to move through 12 degrees, which translates to between 50 minutes and 2.5 hours, depending on how close to the Equator the telescope is located, but it seldom takes more than an hour to make a particular observation.

Michael Cherry

Japan seeks targeted science funding ...

PRIME MINISTER'S OFFICE

[TOKYO] Japan's science policy-makers are planning to make a statement of their priorities in science and technology in an attempt to safeguard funding for key research areas in government budgets. The move marks a departure from the government's traditional approach of spreading research grants thinly and evenly, and would help to identify research areas covered by different ministries and agencies.

The content of the statement — based on a recent survey of science-related ministries and agencies — is currently being discussed by the policy committee of the Council for Science and Technology, the principal science policy-making body, which is chaired by the prime minister.

The policy committee aims to define a number of high-priority research fields that would receive a larger share of government research funds. These will probably include information sciences, genome research, and some areas that were not covered by the five-year plan for science and technology, which was launched in 1996, to double Japan's science spending to a total of ¥17 trillion (US\$140 billion) by 2001.

The statement would be incorporated into guidelines on promoting science, and will be compiled by the end of this month, before ministries start drawing up their budget requests for the next fiscal year. According to the council, the statement will be reflected in the 'special promotion' funds in the budget for the fiscal year 2000.



Looking ahead: Japan's administrative reform committee (above) will determine the future organization of science advice.

Japanese researchers have often criticized the government's inability to adequately support high-priority research areas. For example, many blame Japan's poor track record in genome sequencing on the government's failure to tackle genome research on a national level (see *Nature* 399, 96; 1999). The new plan is intended to distribute research funds more effectively and to explore emerging research areas that need greater support.

According to Nobuhiro Muroya, deputy director of the planning and evaluation division of the Science and Technology Agency (STA), the statement would form an important basis for the second phase of the science and technology basic plan, which begins in April 2001 (see *Nature* 397, 638; 1999).

But he describes the attempt to specify the priorities in scientific research as "hardly straightforward". He adds: "Unsurprisingly, there is strong resistance from areas which

have been dropped from the priority list."

The policy committee also faces a dilemma over how much its new policy should be reflected in future plans, given the impending restructuring of the council in 2001. This will transfer the council to the Cabinet Office, as part of the government's plan to improve the country's administration.

At present, the Science Council relies heavily on STA's administration. The changes would give it its own staff, enabling it to act as an independent body to coordinate all government science and technology.

Some committee members feel that important policy decisions should be postponed until the new council, which will have more authority and an active role in policy making, is formed in two years' time. But others emphasize that the planned statement should be considered as a long-term strategy, saying that the basic policies are unlikely to change under the new council.

"If we are to make a clear statement of priorities in science and technology research, this should become a permanent feature in the policy, rather than a one-off event for the next fiscal year," said Hiroo Imura, former president of Kyoto University and chairman of the policy committee, at the committee's meeting last month.

But Imura admits that there is a limit to what can be decided under the domain of the existing council. "We hope that the new plan would at least be a step up from our previous policy," he says.

Asako Saegusa

... and to use external scrutiny to increase competitiveness at universities

[TOKYO] Japan's Science and Technology Council has called for significant changes in the way university research is carried out. In particular, it wants to create an independent body that would evaluate research externally so as to make it more competitive.

Such a body, according to an interim report on Japan's university research, is needed to compare the quality of research at national universities. The council's report, released at the end of last month, says the results of external evaluations should be reflected in the allocation of government research funds, particularly important given increasing budgetary pressures.

Also included in the report is a plan to create a system by which universities can claim for overheads on research grants.

This would allow researchers to use the grants to cover maintenance costs — which take up a substantial share of government research funds — as basic research funds, which are split equally among universities, decline.

The shift towards competitive and peer-reviewed grant schemes is in line with the government's efforts to make Japanese researchers internationally competitive. At present, according to the council, 88 per cent of Japanese universities, and all but one of its 98 national universities, carry out some form of research evaluation, but only a handful use external reviewers (see *Nature* 397, 378; 1999), most preferring a less stringent self-evaluation system.

Akito Arima, the education

minister and director-general of the Science and Technology Agency, had promised to change the situation by setting up an independent body to assess national universities. His decision came in response to the government's proposal to turn national universities into semi-autonomous institutions, a move strongly resisted both by universities and by the Ministry of Education, Science, Sports and Culture (Monbusho).

Many researchers support the idea of an external evaluation system, but some are concerned that Monbusho lacks the capacity and experience for this.

"The main concern is whether Monbusho's evaluation criteria would be suitable and relevant for the research fields concerned," says Robert Geller, associate professor in

geophysics at Tokyo University.

Geller was recently instrumental in setting up an external review system for reorganizing the university's four departments into a single department of Earth and planetary physics. The review, the fifth to be carried out at Tokyo University, was conducted by 11 external reviewers, five of whom were from outside Japan.

Geller admits that the review involved an "overwhelming amount of work". But he says it helped to identify the strengths and weaknesses of each faculty.

"I am hoping that Monbusho is prepared to create a research evaluation system that rates university departments in a given field on a common scale. If they fail to get it right, it would only be an additional burden to university researchers," he says.

A.S.

US jury split over hormone patent case ...

[SAN DIEGO] A federal jury in San Francisco last week failed to reach a unanimous agreement on whether Genentech Inc. had infringed a patent held by the University of California for DNA for human growth hormone, although it upheld the validity of the patent itself.

After an eight-week trial, eight of the nine jurors agreed that Genentech had infringed the university's patent (see *Nature* **399**, 289 & 297–298; 1999). But the required unanimous decision was blocked by a lone juror.

A further hearing will now be held in two weeks' time in federal court in San Francisco on the next phase of the case, in which university attorneys will seek a new patent infringement trial, while Genentech attorneys plan to attack the university's patent by alleging that it was obtained fraudulently.

The university seeks \$400 million in damages, which it wants tripled to \$1.2 billion because of Genentech's alleged wilful infringement. University attorneys estimate that a new trial will cost \$2–3 million. The university's legal bills already top \$20 million, but those of Genentech are said to be

considerably higher than this.

The trial has drawn particular attention in the scientific community because of dramatic testimony regarding the questionable removal of a DNA sample from a lab at the University of California at San Francisco (UCSF), conflicting testimony by leading molecular biologists, and questions about the veracity of statements in a key paper published in *Nature* in 1979 (*Nature* **281**, 544–548; 1979).

In its patent infringement lawsuit filed in 1990, the university alleged that Genentech used growth hormone DNA misappropriated in 1978 from a UCSF lab to create the company's bacterial synthesis process that produced its growth hormone drugs. Genentech continues to deny the allegations, insisting that it cloned and sequenced the human growth hormone for its drug independently without using the university's DNA.

In an interview after the trial, the jury foreman, Ronald J. Losch, said the eight-member majority of the jury strongly felt that the university had proved its case. "Going through the evidence and applying the law, the conclusion was inescapable," said Losch, a

San Francisco attorney with an engineering background. In deciding the case, Losch said the jury sought to concentrate on the evidence, not to get drawn into a debate over whether one scientist or another was lying.

During the trial, former Genentech researcher Peter H. Seeburg testified that, after Genentech scientists had been unable to clone the DNA for human growth hormone, they used the university's DNA, which Seeburg had spirited out of his former UCSF lab after moving to Genentech. Seeburg also testified that the DNA sample referenced in the paper "didn't exist" — testimony that has prompted him to come under investigation for scientific misconduct at the institute he now directs in Germany (see box).

But Seeburg's claims have been sharply contested by his co-authors, led by David V. Goeddel. A crucial aspect of the evidence involved whether the notebooks of Genentech scientists from 1979 showed a genetic fingerprint of independently cloned growth hormone DNA. Genentech officials and witnesses insisted that they did, while the university's scientific experts testified that they did not.

According to Losch, most of the jury — who, along with the judge, received scientific tutorials during the six weeks of the trial — believed that the notebooks did not provide evidence of a genetic fingerprint. But Goeddel says Genentech's scientific advisory board had examined the notebooks after the trial and felt they did show a fingerprint. "A group of scientists like that is more important than a jury," says Goeddel, now chief executive of Tularik, a company seeking genetic drugs.

As the jury began six days of deliberations, Genentech made its case public by putting parts of the notebooks on a website (see www.gene.com/labnotebooks/). Shortly afterwards, the San Francisco law firm for the university, Morrison & Foerster, augmented its website to include the trial testimony of key witnesses and documents (see www.mofo.com).

Within hours of the decision in the case, Roche Holding announced that it was exercising its option to purchase the final one-third of Genentech's stock, under a previous pact in which Roche bought two-thirds of the stock. Genentech officials said the timing of Roche's move was a coincidence.

At the forthcoming hearing, university attorneys are to ask for a new infringement trial regarding both of Genentech's growth hormone drugs. The university also has another lawsuit pending against Genentech, accusing the company of taking some of the market from an Eli Lilly & Co. drug created from the university's patented growth hormone DNA. Attorneys for Genentech say they may appeal against the jury's verdict validating the university's patent. **Rex Dalton**

... as Seeburg faces misconduct inquiry

[MUNICH] Germany's Max Planck Society (MPS) has begun an investigation into whether Peter Seeburg, a director at the Max Planck Institute for Medical Research in Heidelberg, is guilty of scientific misconduct following his courtroom claim that he misrepresented data in a paper published in *Nature* in 1979 with colleagues from the US biotechnology company Genentech (see above).

The investigation was initiated by Klaus Hahlbrock, vice-president of the society's biology section, after Seeburg wrote to *Nature* and *Science* arguing that the paper had misrepresented experiments involved in cloning the gene responsible for producing human growth hormone.

It will be the first formal investigation to be carried out under new MPS rules for handling cases of alleged scientific misconduct introduced 18 months ago (see *Nature* **390**, 430; 1997).

The investigation



Seeburg: test case for new German misconduct rules.

committee will be chaired by Walter Ordersky, a former president of the federal court. Hahlbrock, three 'arbitrators' from different MPS sections, and the head of the MPS personnel and law department will serve on the committee, which will also call in expert witnesses.

Ordersky will report to the MPS senate "as soon as possible", and recommend sanctions, which may range from a warning to dismissal.

The MPS stresses that the investigation will not address the question of patent infringement or theft of scientific material, the

subjects of the court case that ended last week.

But top-ranking MPS scientists, including Hahlbrock, sensitized by a string of scientific misconduct cases in Germany, say that, had they read a book published in 1987 which described the theft, they would have asked more questions before offering Seeburg a directorship at the Institute for Medical Research in 1996.

Seeburg's position as administrative director — which rotates between the institute's four directors — has been transferred to biophysicist Ken Holmes during the investigation. "It is a strain on the institute, and we are all a bit shell-shocked," says Holmes. But "business is going on as usual," he stresses, with Seeburg's group of 20 or so scientists continuing to collaborate with the group led by Nobel prizewinning electrophysiologist Bert Sakmann, also a director at the institute. **Alison Abbott**

Astronomers win satellite phone curb...

[MUNICH] Europe's radioastronomers have won significant concessions in their prolonged battle with the global mobile-telephone company Iridium over time-sharing in an astronomically important radio waveband.

Under an agreement signed last week, the company will provide 'quiet time' — defined by the International Telecommunications Union as pollution "below the level of detrimental interference" — for facilities in France, Germany, the Netherlands and Britain for seven hours every night and two weekend days per month. The same level of quiet time will be made available to Italy, Poland, Spain and Sweden on request. Interferometry facilities in Britain and the Netherlands will have quiet time every weekend because of their need for continuous 12-hour measurements to map celestial sources.

But the radioastronomers, who joined forces to negotiate through the Committee on Radio Astronomy Frequencies (CRAF), a body associated with the European Science Foundation, are far from satisfied with their victory. They complain that the agreement does not cover legitimate time-sharing, but time-sharing with unnecessary overspill — or 'polluting' — interference.

They argue that Iridium designed its satellite system without regard to its polluting effect, and they are now preparing to fight for a legal limit on overspill at the World Radio Conference later this year.

Iridium's armada of 66 satellites launched last year uses both 'uplinks' and 'downlinks' adjacent to the 1,612-megahertz band that is reserved by international law for radioastronomers. Overspill from the downlinks into this band drowns out weak deep-space signals.

European governments had withheld operating licences from Iridium until the company had reached an agreement with radioastronomers that allowed their research to continue at a "reasonable" level. Last August, an agreement was signed under which the company is to replace all its satellites with non-polluting ones by January 2006 (see *Nature* 394, 607; 1998).

"We are generally happy with our success," says Titus Spoelstra from the Westerbork Observatory in the Netherlands, who is CRAF's frequency manager. He says CRAF was helped enormously by the refusal of individual observatories — unlike those in the United States — to be drawn into isolated negotiations with Iridium, and by the support of European governments. But radioastronomers are not happy to share any time with 'radio waste', he adds.

Across the Atlantic, CRAF's achievements have been highly praised. Mark McKinnon, associate scientist at the National Radio



Time to observe: Italy's Noto radiotelescope is one of those that will benefit from 'quiet time'.

Astronomy Observatory in Greenbank, West Virginia, which has been allocated only four hours' quiet time a night, says this is the most generous agreement made with Iridium.

Mike Davis, spectrum allocation manager at the Arecibo Observatory in Puerto Rico, which negotiated eight hours' quiet time a night, describes the clauses obliging Iridium to stop polluting the 1,612-MHz band altogether after 2006, and concerning weekend observing time, as important victories. Arecibo fought hard for weekend days, but lost, "and this has a great impact on our scheduling [of observing time]."

The most astronomically interesting parts of the Galactic disk pass over Arecibo at fixed

times that vary according to season, and the 1,612-MHz observations are restricted to between May and August, when such regions transit over Arecibo during the night.

At present, neither observatory is having difficulty accommodating the observing proposals, and neither is therefore considering renegotiating its agreement in the light of the European agreement. "But if there is increased pressure of proposals we might think again," says Davis.

CRAF is now turning its attention to the "catalogue of problems that lie ahead", says its chairman, Jim Cohen, a radioastronomer at Britain's Jodrell Bank Observatory. These include threatened pollution in other important radio frequencies, particularly the 1,400–1,427-MHz band which corresponds to emission from atomic hydrogen.

CRAF will be fighting on different fronts for more security over its own radio frequencies at the World Radio Conference, a regular event run by the International Telecommunication Union, the United Nations body that carves up the radio spectrum.

The committee will press for a ban on uncontrolled overspill into wavebands that are allocated to another user. It will also press for clear rules on the allocation of the highest range of the radio spectrum — above 70 GHz — in which commercial companies had, until now, little interest because of the expense of the necessary technology (see *Nature* 390, 103–104; 1997). But commercial pressure is mounting — just as the interest of this millimetre-wavelength range is increasing for astronomers.

Alison Abbott

... and plan for talks on sharing the spectrum

[MUNICH] Research ministers of member states of the Organization for Economic Cooperation and Development (OECD) are to be asked later this month to set up a task force to draw up stricter rules for sharing the radio spectrum between astronomers and telecommunications industries.

The international task force, whose members will include representatives of both sides, will also be asked to find ways of preserving 'radio-quiet zones' for the most sensitive radioastronomy research (see above).

A recently published OECD report points out that

interference in frequencies of importance to radioastronomers will become a much greater problem in the next decade. The report was prepared by a working group set up two years ago by the OECD's Megascience Forum (see *Nature* 390, 103–104; 1997).

Next-generation radiotelescopes are being designed to pick up very weak signals at sub-millimetre wavelengths to allow the study of events very early in the history of the Universe.

But the increased sensitivity of such telescopes makes them proportionately more sensitive to interference, says the report.

Also, commercial pressure from telecommunications industries on high-frequency sub-millimetre bands is set to mushroom.

Satellite records show that there are still a few radio-quiet zones on Earth, for example in western Australia and northern Chile. "We have to keep them quiet," says Harvey Butcher, chairman of the task force and general director of the Netherlands Foundation for Research in Astronomy.

On a task force to plan how to accommodate everyone's interests, Butcher says, "The problem is too big to solve in the long term without everyone getting together".

A.A.

More power sought for US office that protects research subjects

[WASHINGTON] The office that enforces the protection of animal and human subjects in research funded by the Department of Health and Human Services (DHHS) should be elevated to a more visible and powerful position in the office of the DHHS secretary, an expert panel recommended last week (see *Nature* 397, 550; 1999).

Advisers to Harold Varmus, the director of the National Institutes of Health (NIH), said in a report that placing the Office for Protection from Research Risks (OPRR) — currently located in the office of the NIH director — in the charge of a senior government official would obviate any appearance of a conflict of interest and increase the OPRR's influence with other government agencies. Varmus said he would immediately forward the recommendations to Donna Shalala, the DHHS secretary.

Germany to back public understanding of science

[MUNICH] **Germany's Stifterverband für die Deutsche Wissenschaft — the association of foundations that promote the sciences and**

humanities — has launched a DM500,000 (US\$260,000) programme called the Public Understanding of Sciences and Humanities, modelled on a similar scheme in Britain.

The move represents the first step by a group of research organizations, including the Max Planck Society, the Conference of University Rectors and the Deutsche Forschungsgemeinschaft, to create incentives for scientists to convey the significance of research results to the public. An independent jury will award up to DM50,000 to scientists for projects that successfully demonstrate the "immediate use of science and technology for everyday life".

NSF boosts count of US science workers

[WASHINGTON] The US science and technology workforce should be counted at 6–8 million, rather than the 3.2 million commonly cited as holding jobs in traditional science and engineering occupations, according to a report from the National Science Foundation (NSF).

The traditional count includes computer and mathematical scientists, life scientists, physical scientists, social scientists and engineers. But the new figure includes professionals whose jobs regularly call upon their science or engineering education, such

as high-school chemistry teachers or engineers who run manufacturing plants.

Japan says yes to the contraceptive pill

[TOKYO] The Japanese government last week approved the use of the oral contraceptive pill, putting an end to the lengthy debate over its release since the first application was submitted to the Ministry of Health and Welfare nine years ago.

According to the Central Pharmaceutical Affairs Council, an advisory body to the health minister, 16 types of pill, produced by nine drug companies, will be available from September. The panel has also released new guidelines setting up a 10-year post-marketing review period for the pill.

In the past, the government repeatedly declined to release the pill, citing concerns over its potential health risks, including claims that it will lead to an increase in AIDS.

Committee on genetic testing announced

[WASHINGTON] The membership of a new committee to advise Donna Shalala, the US Secretary of Health and Human Services, on genetic testing issues was announced last week. The 13 members of the Secretary's

Advisory Committee on Genetic Testing were selected from almost 200 nominees.

The committee, which first meets on 30 June, will be chaired by Edward McCabe, chairman of the department of paediatrics at the University of California, Los Angeles and an authority on medical and biochemical genetics. It will advise Shalala on all aspects of the development and use of genetic tests, including the medical, ethical, legal and social issues they raise.

Russian PM takes on responsibility for science

[MOSCOW] Russia's new prime minister, Sergei Stepashin, announced last week that he will personally take on responsibility for science. This is in contrast to previous administrations in which the responsibility lay with the deputy prime minister, most recently Vladimir Bulgak (see *Nature* 399, 401; 1999). Stepashin's announcement was made in an address at the Russian Academy of Sciences' 275th anniversary celebrations.

Stepashin told the academy that "the main priority of the cabinet will be to support and collaborate with science and the academic world". Nikolai Platé, the academy's general scientific secretary, said that last year had been "financially, the worst for our academy in the last ten years". At a later meeting,

President Boris Yeltsin said that he intended "to raise the prestige of scientific labour, to make it equal to the highest state services".

Blair advocates 'open mind' on GM food

[LONDON] Britain's prime minister, Tony Blair, appears to have softened his enthusiasm for the commercialization of genetically modified (GM) crops. Speaking on television last weekend, Blair denied that the government was a strong advocate of GM technology. "I'm not the advocate of anything other than keeping an open mind," he said.

His comments appear to have been aimed at defusing controversy stirred up by Prince Charles, the heir to the British throne, who challenged government assurances about the safety of GM food in a two-page newspaper article. But the government's problems worsened this week when the farmer involved in the first commercial-scale experimental trials of GM oilseed rape destroyed the crop on the instructions of the farm's trustees.

Yarden wins first Israeli 'genius' award

[JERUSALEM] Yosef Yarden, a biochemist and dean of the Faculty of Biology at the

Weizmann Institute of Science in Israel, is one of two recipients of a new 'genius' grant for Israeli scholars. The Michael Bruno Awards are to be granted annually by Yad Hanadiv (also known as the Rothschild Foundation) to outstanding Israelis in the field of science and learning.

The award, worth more than US\$200,000, is a no-strings-attached grant intended to free its recipients from other obligations and allow them to devote three years to study and research. Yarden's research has been on how growth hormones govern cell division and the relation of this process to cancer.

UK scientists to work on US fusion facility

[WASHINGTON] The United States and Britain are expected to announce a collaborative arrangement this week that will allow British scientists to work at the new National Ignition Facility (NIF), currently under construction at the Lawrence Livermore National Laboratory in California.

Bill Richardson, the US energy secretary, is due to announce the agreement tomorrow (11 June) at a ceremony being held to 'top out' construction of the building that will house the NIF. A similar agreement already allows French scientists to work with US colleagues on inertial confinement fusion.



The full text of items on this page – and further details and background information about the Unesco/ICSU World Conference on Science, to be held in Budapest from 26 June to 1 July – can be found on *Nature's* website at <http://www.nature.com>

Conference wins the support of prominent politicians

[LONDON] More than 100 ministers of science or education are expected to address the plenary session of the World Conference on Science in Budapest at the end of June, according to Unesco officials.

Their presence at what some officials say might be considered as “the first real world conference on science” is being taken as a sign that many countries are prepared to make a political endorsement of the importance of the conference. They are also seen as supporting its primary theme — the need to establish a ‘new social contract for science’ appropriate to the twenty-first century.

“The idea is that, during the plenary session, governments will make a commitment to science,” said Howard Moore, the secretary of the conference. “We hope this will ensure a real increase in scientific effort.”

One participant will be the German minister of science, Edelgard Bulmahn, who is said to be keen to highlight the situation of women in science, a topic on which she has obtained a commitment to greater activity by Europe’s research ministers.

Another prominent minister who will be speaking is Akito Arima, education and research minister of Japan. Neal Lane, President Bill Clinton’s science adviser, will address the opening plenary session on “the responsibility of the scientist to digest the growing volume of knowledge, to share knowledge with society, and to understand where this new knowledge will lead”.

One science minister who will be missing is Britain’s David Sainsbury. The British paper will probably be delivered by George Foulkes, parliamentary under-secretary of state in the Department for International Development.

Full text: <http://helix.nature.com/wcs/a40.html>

Science academies ‘should encourage ethics debates’

[LONDON] Every academy of science throughout the world should set up a panel of members to study and monitor ethical problems arising from modern science, according to medical physicist Sir Joseph Rotblat, the Nobel peace prizewinner in 1995.

He also suggests that all scientists should be encouraged to sign an oath on graduation pledging to work for the good of humanity. Rotblat is a founder and was for many years president of the Pugwash conferences on science and world affairs, an organization of researchers dedicated to the cause of peace.

“Ethical problems should become part of the terms of reference of all academies of science,” Rotblat told a meeting in London last week. He outlined some of the ideas that he intends to present during a keynote speech on science and human values on the first day of the World Conference on Science.

Acknowledging that many academies are already engaged in such debates, Rotblat suggested that this role should become “normative” rather than optional. The job of scientists nominated to sit on such ethical committees, he said, “will be to try to work out



Rotblat: ‘work for good of humanity’.

the possible outcomes of scientific projects from the point of view of their potential impact on society”.

The physicist pointed out that medical projects involving humans have to be vetted by ethical committees. “Something similar should be carried out for other scientific projects, for example in genetic engineering.”

The code of conduct that Rotblat suggests should be followed by scientists would be similar to the Hippocratic oath taken by doctors. “This makes them aware that their prime duty is to the patient,” he said. “Now that science can affect the fate of so many humans, it is important for scientists to be aware of the situation.”

In his talk, Rotblat will address “the education of scientists as responsible citizens and contributors to the culture of peace”.

Full text: <http://helix.nature.com/wcs/a42.html>

Australian nominated for Unesco top job

[SYDNEY] Australia’s conservative Coalition government announced on Monday (7 June) that it is nominating one of its main political opponents, Gareth Evans, to become director general of Unesco when the post becomes vacant later this year. The surprise move pitches Evans against a lengthy list of nominations, including that from Japan, Koichiro Matsuura (see *Nature* 398, 554; 1999).

Evans, aged 54, has wide international experience. He brokered the Cambodian peace settlement, and was once touted as a potential secretary general of the United Nations. He served as foreign minister in Australia’s Labor government until 1996 and was deputy leader of the opposition until the October 1998 election, after which he did not stand for office.

Full text: <http://helix.nature.com/wcs/a39.html>

Ministers set to back promotion of science in Muslim countries

[LONDON] Twenty science ministers from Muslim countries are expected to give their blessing at the World Conference on Science to plans for a major exhibition on science, technology and medicine in Islam.

The exhibition is to be organized jointly by Unesco and its counterpart organization, the Islamic Scientific, Educational, and Cultural Organization, based in Morocco.

The exhibition will move to a new country every six months. It will begin in Granada, Spain, from where it is expected to follow a route that will take in the Middle

East, North Africa, Sub-Saharan Africa, Asia, Europe and North America.

The ministers will also meet at the conference to agree a strategy document on the future of science in the Muslim world.

Atta-ur-Rahman, coordinator-general of Comstech, the ministerial standing committee for scientific and technological cooperation of the Organization of Islamic Countries, says three critical issues that many developing countries need to solve are skills shortages, excessive bureaucracy, and poor salaries for scientists.

Rahman, who was announced last week as the recipient of this year’s Unesco science prize, believes that governments should pick out “the brightest students from high schools and have them trained in selected disciplines at the best universities and research centres in the Western world”.

Rahman says he believes that the scientific progress of the developing world “will lead to a decrease in international tensions, and lead to a more harmonious and healthier global environment”.

Full text: <http://helix.nature.com/wcs/a41.html>

This Briefing has been written by
Tony Reichhardt

It's sink or swim as a tidal wave of data approaches

Enormous amounts of data are being amassed in fields as diverse as genomics and astronomy. If this information is to be used effectively to speed the pace of discovery, scientists need new ways of working. This requires investment in computers, new statistical tools, and a liberal approach to data sharing.

how demonstrated scientific productivity. But now they are worried. "People recognize there's a problem," says Robert Grossman of the University of Illinois at Chicago's National Center for Data Mining. He has a favourite graph showing the amount of data increasing by more than five orders of magnitude since the early 1980s, while the number of statisticians to analyse the data has remained constant (see page 520).

Are scientists ready for the flood? High-energy physicists, astronomers, climate modellers and others accustomed to supercomputers and high-throughput experiments may be able to cope — given ever more powerful machines, smarter analysis software, and faster networks for gaining access to the databases. But many biologists are still in denial, never having faced the amount of information now pouring into databases such as GenBank and SwissProt.

"Biologists haven't really thought about how they're going to use this data — and the fact that it's going to take real money to do it," says Sylvia Spengler, co-director of the Lawrence Berkeley National Laboratory's Center for Bioinformatics and Computational Genomics in California.

Aravinda Chakravarti of Case Western Reserve University in Ohio, who has been involved in strategic planning for the Human Genome Project, agrees. "The problem is very, very real. Most biologists have no idea of the informatics needs of their projects and how this is likely to change. They are largely unprepared, and the National Institutes of Health has not done enough."

To extract the most knowledge from the vast amounts of raw data now piling up in archives around the world, biologists will need to invest in upgraded computers, more training, and more computer-savvy staff. And they will quickly learn that computer science postdocs are pricier than biology postdocs. The unhappy alternative, warns Spengler, will be to turn over the data to some outsider with the necessary computational skills. In the 'post-genomic' era, it will be a case of learn to work the database, or wait on the sidelines.

Rising to the exabyte challenge

The technical hurdles associated with massive data archives — for biology, physics or any other discipline — may ultimately prove easier to solve than the sociological ones. But they are still daunting. CERN's Large Hadron Collider (LHC), for example, will produce copious amounts of energy and trajectory data for billions of particle collisions every year, for at least 15 years. So important did CERN and its collaborators consider the



Weathering the storm: millions of meteorological records are being transferred from paper to digital storage at the US National Climatic Data Center in Asheville, North Carolina.

Three years ago — several lifetimes in the Information Age — the Microsoft Corporation went in search of a database that it could use to test web-based technology for serving up large amounts of information to the public in a user-friendly way. The bigger and more interesting the archive the better. The New York Stock Exchange had a terabyte (trillion bytes) or so of transaction data, the Mormon church just over two terabytes of genealogical records, but neither of these quite fit the bill.

Then Microsoft approached the US Geological Survey (USGS), which is midway through photographing the entire United States from the air at high resolution. The resulting store of images constitutes about 9 terabytes of uncompressed data, but will grow to 12 terabytes, more than the total amount of printed material in the US Library of Congress. Satisfied, Microsoft signed a cooperative agreement with the USGS, and last year went online with its Terraserver, now among the largest databases accessible on the web.

Such boasts are short-lived in a time when scientists are dreaming up mega-

projects with ever more prodigious output. The US space agency NASA's Earth Observing System will crank out a terabyte of satellite data every day, and a petabyte (1,000 terabytes) every three years. The Large Hadron Collider at the European Laboratory for Particle Physics (CERN) in Switzerland will spew numbers at 20 times that rate.

Information from the collider will flow as fast as if everyone on Earth were each talking into 20 telephones simultaneously. In 15 years it will produce 100 petabytes of stored data — equivalent to generating the total printed content of all US academic research libraries every four months.

Database experts are already bracing for what they call the 'exabyte (1,000 petabyte) challenge'. All the words ever spoken by human beings amount to about five exabytes. In case we need them, and we probably will soon, terms have already been coined for 1,000 exabytes (zettabyte) and 1,000 zettabytes (yottabyte, or 10^{24} bytes).

Managers of large projects such as the Earth Observing System used to brag about such eye-popping numbers as if they some-

On other pages | [Rescuing data: p518](#) | [Cataloguing biodiversity: p519](#) | [Astronomers master digital tools: p520](#) |

Data rescue fills in the climate record

Scientists around the world have moved aggressively in recent years to identify and 'rescue' valuable stores of data — particularly atmospheric and oceanographic records extending back a century or more — before they are lost to the ravages of time.

The records range from handwritten nineteenth-century ships' logs, to weather observations from colonial Africa, to decaying magnetic tape from early weather satellites. The treasures still turn up regularly on dusty shelves and in basements, and even more could be salvaged if a greater public investment were made.

"Every few months we hear about a major archive," says Sydney Levitus of the US National Oceanographic Data Center in Maryland. As an example he cites the Scripps Institution of Oceanography's discovery of 180,000 ocean temperature profiles taken by the US Navy during the Second World War.

No one knew these data existed. But they neatly filled a gap in the North Pacific for which there had been no contemporary record. Altogether, some two million ocean temperature profiles have been added to the centre's World Ocean Database in the past five years, nearly doubling the previous amount.

Part of the credit goes to a campaign to digitize archives already known to exist. But some credit goes to the six-year-old Global Ocean Data Archaeology and Rescue (GODAR) programme, which Levitus heads and which is conducted in association with the Intergovernmental Oceanographic Commission.

GODAR partners meet annually to report on significant data archives in their countries that might be worth saving and converting to digital form for easy electronic sharing. "There's been incredible international cooperation," says Levitus. The Russian Navy has released declassified ocean data, as have the US and British navies.

Most of this historical information, which was gathered for operational purposes, "has never been touched" by scientists, says Thomas Karl, chief of the US National Climatic Data Center. More than 100,000 chlorophyll profiles and 600,000 plankton observations have been added to the database, along with measurements of salinity and ocean chemistry.

The addition of the temperature profiles has already had an important scientific impact, says Levitus. "These data taken together allow us for the first time to compute the interannual variability of the heat storage in the upper ocean," a key factor in climate-change research.

The US National Oceanographic and

Atmospheric Administration has its own data rescue programme. It aims to preserve and eventually digitize millions of old paper, film and tape records stored at three data centres. The task is enormous, and the waiting list is long.

With current funding of \$5 million a year, the agency reckons it will take 18 years to rescue 137 million paper records, and 42 years to salvage 450 terabytes of environmental data stored on outdated computer cartridges and disks. The programme has made great strides, says Karl, "but a lot more could be done".

Ironically, records from the computer age are among the most perishable, with storage media fast becoming outdated. Geostationary Operational Environmental Satellite (GOES) weather data from the 1970s are housed at the University of Wisconsin, where the late meteorologist Verner Suomi had the foresight to transfer them to tape.

But the tapes are deteriorating, and the number of machines that can read them is dwindling. The high-resolution GOES data could yield "potential treasure," says Karl. But the tapes remain in storage because of a lack of money to convert them.

Spurred in part by the signing of the Kyoto treaty on carbon dioxide reductions, scientific organizations around the world have grasped the importance of tracking down and preserving as much climate-related data as possible, says Levitus. "The consciousness has really been raised."

The Belgian government, working with the World Meteorological Organization, has helped African nations to digitize weather observations from the colonial era. Similar work is under way in Caribbean countries, and the European Commission is funding an effort called MEDAR to retrieve data on the Mediterranean Sea.

Funding is sparse for these ad hoc programmes. But enthusiasm is high. The Australian Oceanographic Data Centre recently advertised in a marine science newsletter in an attempt to root out hidden treasure held by individual scientists. Centre head Ben Searle estimates that 70 to 80 per cent of the marine and coastal data that has been collected in Australia "resides in filing cabinets and on personal computers, and its existence is unknown" to all but the owners.

Karl says the tremendous effort that has gone into data rescue should be a sobering lesson for designers of modern electronic databases, who can expect to have to "migrate" terabytes of data periodically to updated storage media, or risk finding themselves some day with a "dead archive". It won't be glamorous work, he says. But "we'll all have to pay some tax to do this".

data management task that they started working on it in 1995, ten years ahead of the accelerator's coming online.

The LHC's data system will push the state of the art in several areas, including high-performance storage for computer data, where capacity may be less of a problem than speed of transfer between tapes and other media. Fortunately, the commercial world is working on the same problem. LHC data managers therefore hope to buy the storage system and database software essentially off the shelf, then modify them.

The four individual LHC experiments have already made a decision to forsake tried-and-true Fortran code and switch to more versatile object-oriented programming (of which C++ and Java are common varieties). This will require "a different mindset" for CERN programmers, says Julian Bunn, a CERN physicist working at the California Institute of Technology, which is collaborating on the LHC data system.

It was a bold step, not taken lightly. NASA's Earth Observing System data system (EOS-DIS) crashed on the rocks of object-oriented programming in the early 1990s, when the tools were newer. The project suffered delays and cost overruns as a result.

The LHC data system will be distributed, with a central archive at CERN and regional centres serving users nearest to them. Scientists tapping into the database, though, will have the impression of a single repository. Bunn worries most about networking — moving around enormous volumes of data, quickly and seamlessly, among ten regional centres, the central archive at CERN, and up to 2,000 individual users around the world.

A new, higher-capacity Internet should be available by then. But millions of users worldwide will be soaking up the bandwidth by downloading movies and shopping online. Even dedicated scientific networks are likely to fill up quickly, says Bunn.

The life expectancy of the archive is equally problematic. "We've never been faced with maintaining a 20- or 25-year database," says Bunn. As the LHC matures, the calibration of its instruments will become more refined, and all those petabytes of data will have to be regularly reprocessed, adding to the volume of data the system has to churn. Project managers only expect 100 petabytes from the collider itself. But they are designing their system to handle ten times more data.

What's worth keeping?

Computational demands will become greater as scientists build more linkages and capabilities into their databases. Services such as the US National Center for Biotechnology Information have already begun to merge scientific journals and databases into unified searchable libraries (see Briefing, *Nature* 397, 195–200; 1999).

Catalogue of life could become reality

Taxonomists have long dreamt of creating a master 'catalogue of life'. For various reasons — lack of money and competing schemes for going about the job being the two most prominent — it has not yet happened. But the 1992 global biodiversity treaty may be a spur to further action.

Frank Bisby of the Centre for Plant Diversity and Systematics at the University of Reading, England, hopes that the Species 2000 project to federate as many as 200 databases into a single searchable archive of all the world's 1.7 million known species "is about to turn from a plan into a reality".

Other groups with similar ambitions have signed on as partners, including the US-Canadian Integrated Taxonomic Information System and the Global Plant Checklist based in Australia and Europe.

Today, the prototype Species 2000 'dynamic checklist' searches just three databases. But up to 30 links are expected by

the end of the year. Bisby is also discussing links to the geospatial database created by the University of California at Santa Barbara's Alexandria Digital Library, so that species data could be combined with geographic information.

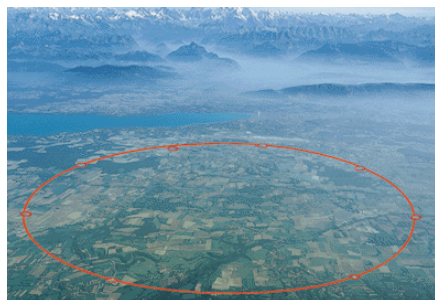
Funding so far has been "ridiculously fragile," he says. But he is optimistic that the Global Environment Facility and the European Union will help pay some of the estimated \$140 million cost of the basic (non-georeferenced) system.

Bisby and other taxonomists have been envious of the ample funding bestowed on molecular biologists, when "we think of ourselves as being of equal stature". Chris Thompson, a US Department of Agriculture researcher and former vice-chairman of Species 2000, says: "We've just been ineffectual at selling our vision." Species 2000 is at www.sp2000.org

A data workshop sponsored by the US National Science Foundation last year produced an even more ambitious vision: digital journals that would link not only to the data used in an experiment, but to the programs that created or analysed the data, so that readers could verify the results of an experiment or run their own variations. The workshop participants called this 'deep citation'.

"That scares me a bit," admits Bunn. He understands the appeal to scientists who want to recompute some controversial result for themselves. But who would be allowed access to the data and the computational resources, and on what terms? It's an idle worry today, he says, but "I guess it will come".

Uncertainty about what information to keep — a particular problem in young fields such as genomics — will be a big contributor to database bloat. Are all expressed sequence tags sent to GenBank worth keeping? And all single nucleotide polymorphisms? Before scientists have thoroughly analysed the data, no one can say. Until then, says Spengler, "you have to be a pack rat" to avoid throwing away something important.



Data factory: handling experimental results from CERN's Large Hadron Collider (above), currently under construction, has presented a daunting challenge to physicists.

It's unwise to cater to every scientist's whim about what data should be archived, says Graham Cameron, joint head of the European Bioinformatics Institute (EBI) outside Cambridge, England, which maintains, among others, the SwissProt and EMBL nucleotide sequence databases. "You could soak up any amount of money" trying to store everything, he says. Database managers have to judge what data are used most often by scientists, and what might be used in the future.

This is not easy. "Complexity fights its way in," says Cameron. The EBI recently decided to establish a public domain repository for DNA microarray-based gene expression data, despite concerns that such an archive might be premature (see *Nature* 398, 646; 1999). Cameron calls this a "strategic" commitment, even though "technically we may not be at the stage yet to do it right".

Maynard Olson, a geneticist at the University of Washington who has been deeply involved in the Human Genome Project, thinks that the rush to produce a 'quick-and-dirty' draft of the human genome may lead to headaches later, as a mountain of low-quality data is harder to analyse than a smaller, more refined dataset.

Who will pay?

With the web firmly established as a primary avenue of scientific communication, the notion of a database as a large repository in a single location has become *passé*. Grossman, of the National Center for Data Mining at Chicago, says "the tide is shifting pretty dramatically to distributed systems," which can be loose federations of independently operated databases using common data standards and transfer protocols.

The federations may not be as efficient as

centrally managed archives — "every link you build sets up a dependency," says Cameron — but they have real advantages, such as allowing specialists to keep and curate their own data.

Concern about ownership remains an obstacle to database sharing. Researchers at the University of Kansas Natural History Museum hope they have found one solution in a data retrieval protocol called Z39.50, which has proven successful in the bibliographic community. A Z39.50 query retrieves and pools data from multiple sources — perhaps museums in different locations that hold specimens from the same taxon or region. Each museum retains control of its own database, but the pooled results add up to something no single collection could offer — enough data points to allow detailed analysis of biodiversity patterns.

But the issue of data ownership will not go away easily, especially for information with perceived commercial value, and this could prevent many scientists from making the most of the current data bonanza. At least some new information on the human genome — particularly products derived from raw sequence data — will be off limits to those who do not pay private companies such as Celera for access rights (Novartis, Upjohn and other large companies have already done so).

Scientists worry that proposed changes to US intellectual property laws could push researchers to view their data as commodities to be sold rather than as information to be shared (see *Nature* 394, 410; 1998).

Cameron at the EBI places some of the blame on stingy governments. SwissProt reluctantly began charging commercial users for access to its database only after government funding dried up. The present situation is "not ideal," he says, but it reflects the "inability of the public funding mechanism in Europe" to recognize the importance of free genomic data.

The commercialization of research databases could also shut poorer developing nations out of the scientific mainstream. But the ramifications go beyond North-South politics. So agitated did European nations become over US companies' practice of gathering free European meteorological data, then repackaging them into commercial databases sold back to Europeans, that the World Meteorological Organization passed a resolution several years ago allowing countries to restrict access to certain kinds of commercially valuable weather data.

Until that time, the information had always been shared freely among nations. "No question about it, it was a step back," says Michael Crowe, science planning officer at the US National Climatic Data Center. Data exchange and intellectual property rights are "becoming a thornier and thornier issue," he says.

The good news in today's data explosion

is that not a day passes without some government agency, academic consortium or private company developing a clever web interface or piece of data analysis software that makes sifting through the petabytes that much easier.

Any number of indexes and 'master directories' of environmental and global-change data have opened online in the past two years — perhaps too many, says Thomas Karl, director of the US National Climatic Data Center. "I know this is heresy for the head of a data centre," he says. But advertising a one-stop shop for climate data is "some data manager's pipedream". Savvy scientists will never buy the idea of a single place that has it all.

Most researchers are accustomed to studying a relatively small data set for a long time, using statistical models to tease out patterns. "At some fundamental level that paradigm has broken down," says Grossman. "You can't be afraid of data today."

Soon the question for scientists will be "how do you manage a terabyte in front of you?" — an even more difficult challenge considering that a computer may need to generate 1,000 times that amount in the course of manipulating the data.

Mining for data

Various schemes have been proposed to extract knowledge from such large stores of information, including supercomputer-powered scientific 'visualization' and what's been called 'data mining', or the semi-automated discovery of patterns, associations and statistically significant structures. Data mining borrows tricks from other fields such as artificial intelligence and neural networks to produce software that can plough through large datasets, looking for nuggets that humans would take forever to find.

Astronomers at the California Institute of Technology used a program called Skicat to automatically sort through three terabytes of image data from the Palomar Observatory Sky Survey. Using decision trees and classification rules, the system was able quickly and accurately to classify very faint objects, effectively tripling the number of objects in the catalogue and accelerating the pace at which high-redshift quasars were discovered. The

Reaching for the digital sky

'All-sky' surveys are a popular undertaking in astronomy these days. The 2MASS project is using telescopes in Arizona and Chile to photograph northern and southern skies in the 2-micrometre infrared range. The Digital Palomar Observatory Sky Survey is expected to catalogue some two billion sources in the north, and the Very Large Array in New Mexico has two sky surveys in progress.

These and other searches will generate tens of terabytes of high-quality astronomical images and data. The goal of the Digital Sky project is to show how all this information might be merged into a "multi-wavelength digital library covering a significant fraction of the real sky".

Astrophysicist Tom Prince of the California Institute of Technology is the principal investigator, with funds coming from the US National Science Foundation's

National Partnerships for Advanced Computational Infrastructure programme. Digital Sky will focus on the "switchyard" problem, says Prince — how to develop standards and techniques for merging data with different formats into a virtual, distributed database. The Sloan Digital Sky Survey, the most ambitious of the planned surveys, will not be incorporated in the Digital Sky database, but will be among the groups cooperating in setting the standards.

The payoff for future operational databases could be enormous. Scientists might, for example, be able to search easily a billion astronomical objects across a wide range of wavelengths, looking for unusual patterns of energy output.

Digital Sky is at www.cacr.caltech.edu/SDA/digital_sky.html

program was developed by a Jet Propulsion Laboratory software engineer (who later moved to Microsoft).

Scientists at the EBI have mined large stores of yeast gene expression data. Astronomers at Australia's Mount Stromlo and Siding Springs Observatories used mining programs to hunt through data on 20 million stars taken nightly for four years, making more efficient the search for Compact Halo Objects (MACHOS).

Data mining is sometimes overhyped, admits Grossman. Some call it "data analysis with better marketing," while others worry that machines searching blindly for subtle associations in large datasets could do "very stupid things" such as correlate heart attacks with the stars. Human judgement will always be necessary, he says. Data mining is just a tool, albeit a powerful one.

Richard Gibbs, a gene sequencer at Baylor College of Medicine in Texas, believes it is imperative in the era of high-throughput genomics for biologists not to turn over all the action to computer scientists, who might miss biologically important information.

Francis Collins, head of the US National Human Genome Research Institute

(NHGRI), agrees. "When I give talks to young scientists seeking advice about areas of future intense scientific excitement, computational biology is my number one recommendation," he says. Money alone will not effect change, although training grants in bioinformatics have been stepped up by federal agencies such as the National Institutes of Health (NIH) and the Department of Energy, and private groups such as the Pharmaceutical Research and Manufacturers of America Foundation.

Lisa Brooks, programme director for genome informatics at the NHGRI, says there are two pressing needs when it comes to training. One is for biologists to learn to use publicly available genome analysis software (developed at NIH and elsewhere) through courses, online tutorials, and mentoring with other scientists.

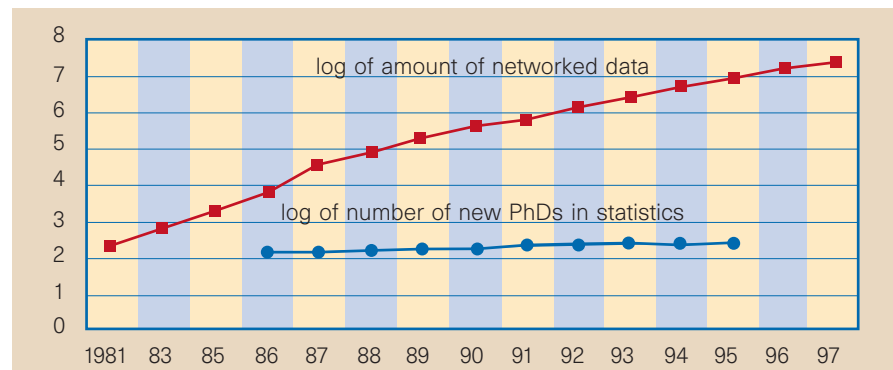
That, she says, is the relatively easy part. Much harder will be building up a corps of scientists who can come up with new statistical approaches to analysing large volumes of genomic data. Today, according to Brooks, "only a small proportion of biologists are capable of developing the tools".

Bioinformatics is still fairly new, and many training programmes are taking only a few students, says Collins. The discipline has had trouble establishing itself.

"Computational biologists in academia often find themselves without a clear career track or an academic home," he says. "Their efforts are seen as too applied to earn respect in departments of computer science, and too ethereal to be accepted in biochemistry or physiology. We need a new mindset."

Gibbs agrees, and says that scientists have no choice but to adapt to a data-rich world. "We've all been doing it slowly, but we've all got to do it faster. The accelerated pace of this is just leaving everybody breathless." □

SOURCE: NCDM



Skills gap: networked data rises, while the number of those trained to handle it remains constant.

Exploitation of junior scientists must end

Sir—There has been a longstanding need for serious analysis of the genuine ethical problems of exploitation, corruption and abuse in science¹. Unfortunately, the debate instead tends to be framed in terms of ‘fabrication, falsification and plagiarism’.

On the one hand, we have seen intense and minute public scrutiny of whether a few individuals may have presented false or misleading scientific information; and on the other, independent juries have confirmed that research has been stolen and that corruption has occurred at universities^{2,3}. Most horrifying of all, Jason Altom, a promising graduate student, felt so pressured and trapped by academic “abuse” at Harvard that he took his life⁴.

So we have an elite society that exploits junior colleagues to the point of suicide, but we are choosing to discuss not what is fundamentally wrong with the system, but whether the system produces accurate information at all times.

Two important points about scientific ethics have not, to my knowledge, been discussed together. First, scientific falsification is both rare and severely punished. This is intrinsic to the nature of science: a falsification that has any meaningful implications will sooner or later be found out as its implications are examined. Likewise scientists are understandably fiercely protective of the perceived value of their product, so punishment for falsification is severe.

Second, exploitation of junior scientists by their seniors has arguably become commonplace, but is rarely discouraged — on the contrary, scientists who exploit their juniors are richly rewarded.

There is considerable anecdotal evidence for these views. For example, a researcher at a large private university was reportedly berated and then struck in the face by her academic supervisor when, after working without pay for two years, she requested a salary. The senior professor in charge of the laboratory then fired the researcher in response to a court judgement against the supervisor⁵.

In another example, according to independent sources, senior professors at different universities routinely hire foreign postdocs who have to work indefinitely under conditions that they stipulate by refusing to write letters of recommendation.

And in a third example, the chairperson of a university department has been accused of changing the budgets on grants written by junior researchers to support himself rather than the grant authors. He then allegedly stopped their salaries without notice and demanded that the researchers ghost-write further grant proposals as the price for reinstatement.

The atmosphere that would permit a professor to think that such behaviour could be acceptable seems to be clear evidence that the system is badly broken. Nor are these incidents obscure. On the

contrary, the first case was reported in campus and city newspapers, and the last was reported at the time to university officials at the highest level. No action was taken against any of the senior faculty members involved. In each case he remains in good official standing.

The strictures against open confrontation of offenders are also well known. A certain method of ending an academic career is to protest openly against mistreatment — however clear the evidence or egregious the offence.

Discussing, in this atmosphere, whether or not fabrication, falsification or plagiarism may occur seems to me like asking whether employees in a company accused of being a sweatshop pilfer thread or make shoddy clothes. It is time for scientists of merit to address the real issues of concern to the next generation of scientists.

Are incidents such as those I report common? Can responsible scientists take measures to protect students and junior researchers? These are the issues crying out for attention, not whether data has ever been pilfered or if research might be shoddy.

Troy Shinbrot

Department of Chemical & Biochemical Engineering, Rutgers University, Piscataway, New Jersey 08854, USA

1. Abbott, A. *et al.* *Nature* **398**, 13–17 (1999).

2. Charatan, F. B. *Br. Med. J.* **315**, 501 (1997).

3. *Science* **265**, 463 (1994).

4. Nadis, S. *Nature* **395**, 826 (1998).

5. Zorn, E. “Chicagoland” p. 1, *Chicago Tribune* (25 Feb. 1993).

Lifting the lid on the homeobox discovery

Sir—In reviewing my book, *Master Control Genes in Development and Evolution*, William McGinnis and Peter A. Lawrence¹ invoke Lewis Carroll’s remark that “your view depends on where you’re standing”. The remark is only witty if the facts are correct. In this review, they are not.

Like many a discovery, that of the homeobox was sparked by a vague indication that could either be disregarded as irrelevant, or followed up. This was a weak band on a gel of Richard Garber’s which most members of my research group considered to be an artefact, but to me this was the first sign of the homeobox. The reviewers claim that “[this band] was attributed to overloading of the gel, lumped into the ‘uninterpretable results’ category and not followed up”. This statement is incorrect. We reproduced this band and documented it in two papers.

The finding of this cross-hybridizing band was first published in Garber’s paper²: “Under stringent hybridization conditions, weak homology with both the 903 and 909 probes was detected at position 190 kb... These findings are being investigated further”; and subsequently in McGinnis’s paper³ describing the discovery of the homeobox: “Garber *et al.* found a weak homology between the 903 cDNA and a site to the left of the *Antp* locus at position 190 on the map in Fig. 1. This site has subsequently been shown to be part of the transcription unit of the *fushi tarazu* gene (A. K. and E. H. in preparation)”.

Would we have published this finding if I had judged it an artefact “lumped into uninterpretable results”? In fact, it was the first sign of the homeobox that caused me great excitement. It was McGinnis who analysed this weak homology cleanly and determined the sequence of the homeobox. Because of his important contribution, he fully deserved to be first author of the paper describing the homeobox. Michael Levine, Atsushi Kuroiwa and Ernst Hafen were

additional key contributors to this important discovery. Why is it so hard for McGinnis and Lawrence to accept that I, too, may have had a share in it?

Walter Gehring

Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

McGinnis and Lawrence reply—Gehring states that there is documentary evidence that it was he alone of his research group who appreciated that a weak band on Garber’s Southern blot of autumn 1982 was the first sign of the homeobox. This evidence consists of the statement in the November 1983 paper of Garber *et al.*² that “weak homology was detected between 903 and 909 probes (Antennapedia cDNAs) and position 190 (of the genomic walk that included the Antennapedia transcription unit)”. However, this paper was submitted on 18 July 1983, after the significance of the homeobox cross-hybridization had become clear to everyone in Gehring’s lab — but by the different route described in our review.

In July 1983, the status of homeobox

research was as follows. A few novel *Drosophila* homeobox clones had been isolated, and one of these was defined as a 3' exon of the *Ultrabithorax* (homeotic) gene. All the novel clones mapped to cytogenetic regions that were known to contain the two homeotic gene clusters in flies. The transcripts encoded in these clones were expressed in unique, homeotic-like stripes on the anteroposterior axis of developing embryos. So, at the time the 1983 Garber paper was written, it was known in Gehring's lab (but only retrospectively) that Garber's band was not an artefact.

It is always possible that Gehring sensed or knew in 1982 that Garber's band was a crucial clue which should be the basis of further investigation. However, to our knowledge no one heard any such suggestion, nothing was done about it, and it did not spark the crucial experiments on the homeobox sequence in *Antennapedia* and other developmental control genes.

William McGinnis

Department of Biology 0349, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0349, USA

Peter Lawrence

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

1. McGinnis, W. & Lawrence, P. A. *Nature* **398**, 301–302 (1999).
2. Garber, R., Kuroiwa, A. & Gehring, W. J. *EMBO J.* **2**, 2027–2036 (1983).
3. McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. *Nature* **308**, 428–433 (1984).

Longevity – does family size matter?

Sir — Rudi Westendorp and Thomas Kirkwood¹ conclude that women who live longer have fewer children. This can be ascribed almost entirely to an increase in the proportion of childless women in higher age groups. For those women that have children, the mean number of children increases gradually, with a maximum in the 71–80-year-old group, followed by a slight downward trend which is not significant (Table 1).

The overrepresentation of childless women in high age groups suggests that

giving birth and raising children shortens life expectancy. But, once you have children, the number you have makes no difference to your life expectancy. Therefore, it is not a matter of reduced fertility, but a case of 'to have or have not'.

Toon Ligtenberg, Henk Brand

Department of Oral Biology, Academic Centre for Dentistry, van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands

Westendorp and Kirkwood reply — We found¹ that aristocratic women had low average family sizes, the reasons for which are discussed by Jim Cummins². In these circumstances, any impairment of fertility is likely to result in an increased likelihood of remaining childless. Excluding childless couples from the analysis is therefore counterintuitive. In our opinion, the data from the British aristocracy do not support Ligtenberg and Brand's conclusion that "once you have children, the number you have makes no difference to your life expectancy".

A study³ of 822,593 women from the Norwegian census of 1970 found that, among post-menopausal women, those with larger numbers of children (more than four) also had higher mortality rates. This is consistent with our finding of a negative association between longevity and reproductive success; furthermore, it suggests that family size does matter.

In his News and Views article about our paper, Daniel Promislow⁴ suggested that environmental, rather than genetic, factors might explain the trade-off between longevity and reproductive success, which we showed was similar for women and for men. For example, a large family might increase environmental stress and mortality risk for both parents. If this was the case, spouses' lifespans should be correlated. We found a statistically significant correlation, but it accounted for only 2% of the variance in age at death. The weakness of this correlation argues strongly against environmental factors playing a major role in the trade-off, and supports the hypothesis that genetic factors are important.

R. G. J. Westendorp*†, T. B. L. Kirkwood†

*Section of Gerontology and Geriatrics, Department of General Internal Medicine, and Clinical Epidemiology, Leiden University Medical Centre CO-

P, PO Box 9600, 2300 RC Leiden, The Netherlands

†Biological Gerontology Group, Department of Geriatric Medicine and The School of Biological Sciences, University of Manchester, 3.239 Stopford Building, Oxford Road, Manchester M13 9PT, UK

1. Westendorp, R. G. J. & Kirkwood, T. B. L. *Nature* **396**, 743–746 (1998).
2. Cummins, J. *Nature* **397**, 557–558 (1999).
3. Lund, E., Arnesen, E. & Borgan, J.-K. *J. Epidemiol. Community Health* **44**, 237–240 (1990).
4. Promislow, D. E. L. *Nature* **396**, 719–720 (1998).

City dwellers must share blame on biodiversity

Sir — In your editorial on the debate over genetically modified foods, you say "the steady reduction of biodiversity remains a silent witness to the potential of modern agriculture to inflict damage on the environment and the wildlife it supports" (*Nature* **398**, 639; 1999). It is unfair to attribute solely and without qualification to agriculture an effect that is correlated with a whole raft of land-use changes and also with climate change.

Many farmers take great care over the countryside. You can also argue that city dwellers contribute to the reduction in biodiversity by demanding more housing development, more motorways, and so on.

A. J. Murdoch

Department of Agriculture, University of Reading, Earley Gate, PO Box 236, Reading RG6 6AT, UK

Biblical answer to cooking up pi

Sir — In the News story about scientists' response to creationists, the scientists "comment that the Bible says that π is 3, not 3.14" (*Nature* **398**, 453; 1999).

The biblical verse quoted (1 Kings 7:23) reads in part: "...measuring 10 cubits from rim to rim... It took a line of 30 cubits to measure around it". Indeed, 30/10 equals 3, but further on in verse 26 it says: "It was a handbreadth in thickness...". Assuming that a cubit measured 18 inches and a handbreadth 3 inches, the inner diameter of the bowl would be 174 inches ($10 \times 18 - 2 \times 3$), and the inner circumference would be 540 inches (30×18). This yields a value for π of 540/174 or 3.10. This is about a 1 per cent error from the typical value for π of 3.14. Although we do not know the exact length of a cubit or a handbreadth, this result is very close to the actual value of π .

Kevin Peil

Reactor Engineering Group, Dow Chemical Company, Midland, Michigan 48674, USA

Table 1 Relationship between age at death and number of children for married aristocratic women

Age at death (years)	Proportion childless	Number of children	
		mean for all women	mean for women having children
<20	0.66	0.45	1.32
21–30	0.39	1.35	2.21
31–40	0.26	2.05	2.77
41–50	0.31	2.01	2.91
51–60	0.28	2.4	3.33
61–70	0.33	2.36	3.52
71–80	0.31	2.64	3.83
81–90	0.45	2.08	3.78
>90	0.49	1.80	3.53

Shifting seas in the greenhouse?

Stefan Rahmstorf

Models of the Earth's possible responses to global warming are continually being improved. The latest simulation of changes in deep flow in the Atlantic operates without several of the fudge factors previously required.

The effect that global warming might have on the circulation of the Atlantic Ocean has been a topic of much speculation and research. On page 572 of this issue, Wood *et al.*¹ present greenhouse warming scenarios computed with a climate model that, for the first time, gives a realistic simulation of the large-scale ocean currents without requiring artificial adjustments of the air–sea fluxes.

Of more immediate interest to those outside the modelling business, Wood and colleagues' results show a dramatic change in the Atlantic occurring over the next few decades: a complete shutdown of one of the two main 'pumps' driving the formation of North Atlantic Deep Water, namely the one in the Labrador Sea (Fig. 1).

In 1987, in an article entitled "Unpleasant surprises in the greenhouse?", Broecker² warned that the response of the climate system to greenhouse warming might involve 'mode switches' of the Atlantic circulation. He drew this inference from palaeoclimatic data, indicating that such events had occurred in the past, and from early ocean modelling results. Initially, the idea was simply that a positive feedback meant that the large-scale overturning motion of the Atlantic (sometimes popularly dubbed the 'conveyor belt', in which warm surface waters flow northwards and cold deep water returns south throughout the Atlantic, acting like a central-heating system for Europe; see Fig. 1) could exist in two distinct states — switched on (as at present) or switched off.

Later work revealed a more complex picture, by showing that individual sites of oceanic convection could also have a tendency towards flip-flop behaviour (switching between quasi-stable states with convection 'on' or 'off')^{3,4}. Given that there are two main sites of convection linked to the formation of North Atlantic Deep Water, in the Greenland Sea and in the Labrador Sea, this led to speculation that global warming could switch off one of these convection sites⁵.

Wood *et al.*¹ provide the first clear modelling result indicating that this could indeed be the response to increasing concentrations of greenhouse gases. The team works at the Hadley Centre in Britain, and their climate model is remarkable in several respects. Cer-

tain improvements, including a higher resolution of the ocean and parameterizations of eddy mixing and of the flow of bottom currents over marine sills, mean that the model provides a more realistic representation of the main ocean currents than previous coupled ocean–atmosphere models. In particular, the partitioning of the deep-water formation between the Greenland and Labrador Seas is in good agreement with observations. Furthermore, unlike most previous climate models, this model does not use or require 'flux adjustments' at the air–sea interface, which involve adding a prescribed heat or freshwater flux to make up for a mismatch between ocean and atmosphere components of the climate model. Flux adjustments had to be used in the past to prevent the model climate from drifting slowly to unrealistic conditions, but they may distort the stability of the ocean circulation⁶.

The authors subject their model to two greenhouse-gas scenarios: an artificially rapid increase in atmospheric concentrations by 2% per year up to a quadrupling of CO₂, and a more realistic 'business as usual' scenario starting in 1860 and extending to the year 2100. In both cases, the overall volume of water transported by the Atlantic conveyor belt (Atlantic overturning) decreases by around 25%, but it does not reach the point of collapse. This is consistent

with most simulations by other groups (Fig. 2, overleaf). Previous studies indicate that the threshold for a complete conveyor-belt shutdown could be crossed in the twenty-second century if precipitation and meltwater runoff into the North Atlantic are strongly enhanced⁷, diluting the surface waters to the point where the high-latitude sinking motion stops. This, however, is a point on which large uncertainty remains in climate models.

What is new in Wood and colleagues' simulation¹ is the shutdown in Labrador Sea convection and the associated collapse of the Labrador Current, which occurs between the years 2000 and 2030. If such a collapse did occur, it could have serious consequences for marine ecosystems, including seabird populations in the region, as they depend not only on specific temperature conditions but also on nutrients supplied by oceanic mixing and currents.

But how likely is this change in ocean circulation? From a single model simulation, even one carried out with an advanced climate model, it is difficult to assess the range of uncertainty. Many regional aspects of the ocean circulation, including the energetic synoptic eddies, cannot at present be resolved by this (or any other) climate model. Synoptic eddies are the ocean's equivalent of the high- and low-pressure systems that make up weather in the atmosphere, and they play an important role in mixing ocean waters and in transporting heat in some regions. High-resolution ocean models exist, but their severe computational cost means that they cannot yet be used in climate simulations. Owing to the lack of relevant controlled experiments comparing high- and low-resolution models, it is still an open scientific question to what extent the results of climate models would change if oceanic eddies were resolved explicitly.

Another caveat is the uncertainty in regional precipitation changes due to global

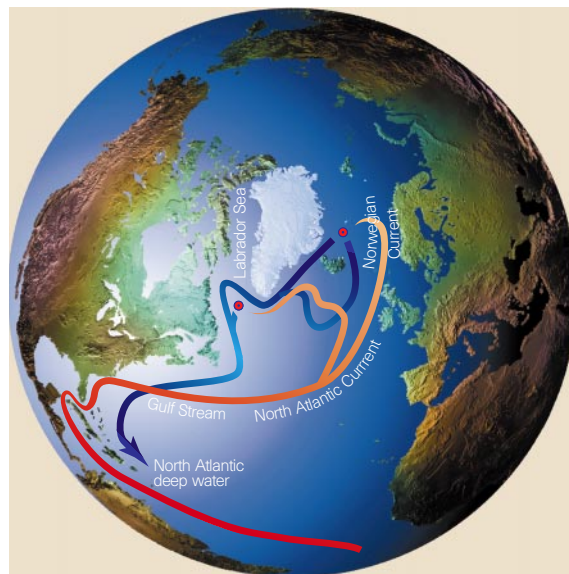


Figure 1 Simplified sketch of currents in the North Atlantic, showing the two main convection sites in the Greenland and Labrador Seas. Warm surface currents are shown in red; cold, deep currents in blue. Red-and-blue circles, convection sites.

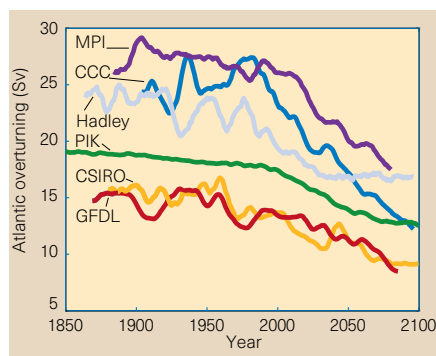


Figure 2 Simulated water-volume transport of the Atlantic 'conveyor belt' (Atlantic overturning) in a range of global warming scenarios computed by different climate research centres. Units are Sverdrups, $10^6 \text{ m}^3 \text{ s}^{-1}$. Although a wide range of transport is simulated for the unperturbed climate, all of the models generally show little change up to the present. But all show a significant weakening of the flow in the next century as a result of greenhouse warming. Data shown are nine-year running mean values. CCC, Canadian Centre for Climate Modelling and Analysis, Victoria¹¹. CSIRO, Division of Atmospheric Research, Melbourne¹². GFDL, R30 model of the Geophysical Fluid Dynamics Laboratory, Princeton (ensemble of three runs)¹³. Hadley, HadCM3 model of the Hadley Centre, Bracknell, UK¹. MPI, ECHAM3/LSG model of the Max Planck Institute for Meteorology, Hamburg¹⁴. PIK, CLIMBER-2 model of the Potsdam Institute for Climate Impact Research⁷.

warming, as the input of fresh water from increased precipitation would have a considerable effect on oceanic convection by changing the surface water density. Different models of atmospheric behaviour predict widely differing changes in precipitation. The Hadley Centre model also does not take into account the fresh water entering the Labrador Sea from the north, through the Canadian archipelago.

Given these uncertainties, the results of Wood *et al.* should be interpreted as a warning that a regional shut-down of convection could occur and that it could even occur soon, without taking the specific timing and location as a definitive prediction. The Labrador convection site may indeed be more vulnerable to greenhouse warming than the Greenland Sea site, given that the deep water formed in the Labrador Sea is already the less dense of the two, which could mean that a relatively small reduction in surface density could stop convection altogether. This hypothesis needs to be confirmed by future modelling studies.

So far, however, observations show no sign of a systematic weakening of convection in the Labrador Sea. On the contrary, after a weak phase in the late 1960s, convection in the 1990s has been favoured by the high phase of the North Atlantic Oscillation⁸, the

dominant mode of natural variability in the region. Both North Atlantic convection regions need to be continuously monitored, as planned in the international CLIVAR (Climate Variability and Predictability) programme.

The simulated ending of Labrador Sea convection found in the Hadley Centre model is perhaps the most convincing demonstration so far of a qualitative threshold being crossed because of global warming. Such nonlinear effects are increasingly receiving attention in climate research, following the 1995 IPCC report's⁹ warning that "future climate change may also involve 'surprises'". Another possible 'surprise' (the term is a misnomer, as the element of surprise lessens the more it is discussed) is that the West Antarctic Ice Sheet could become unstable and slide into the ocean, causing sea level to rise by several metres¹⁰.

To evaluate the risk of such events it is not enough to compute a few 'best guess' projections of future climate change. More sophisticated and systematic approaches to risk assessment are required, taking the full range of uncertainty in current knowledge and models into account. This is a difficult task, not helped by the fact that many climate models operate at the limits of available computer power. Nevertheless, it needs to be tackled. As in any risk assessment, it is not just the most likely outcome of global warming that affects how we should handle the

problem, but also the 'tail ends' of the probability distribution of possible futures: events with a low probability but a high damage potential. Take the risk of nuclear power stations for comparison: the most likely scenario is that everything works just fine, but large investments are nevertheless made to insure against the tiny probability of a catastrophic failure.

Although many global warming risks are still difficult to quantify, the work of Wood and colleagues is a reminder that now is the time to invest in a 'full-cover insurance policy' by reducing greenhouse gas emissions. □

Stefan Rahmstorf is at the Potsdam Institute for Climate Impact Research (PIK), PO Box 60 12 03, 14412 Potsdam, Germany.
e-mail: stefan@pik-potsdam.de

1. Wood, R. A., Keen, A. B., Mitchell, J. F. B. & Gregory, J. M. *Nature* **399**, 572–575 (1999).
2. Broecker, W. *Nature* **328**, 123–126 (1987).
3. Lenderink, G. & Haarsma, R. J. *J. Phys. Oceanogr.* **24**, 1480–1493 (1994).
4. Rahmstorf, S. *Nature* **378**, 145–149 (1995).
5. Rahmstorf, S. *Nature* **388**, 825–826 (1997).
6. Marotzke, J. & Stone, P. H. *J. Phys. Oceanogr.* **25**, 1350–1364 (1995).
7. Rahmstorf, S. & Ganopolski, A. *Clim. Change* (in the press).
8. Dickson, B. *Nature* **386**, 649–650 (1997).
9. Houghton, J. T. *et al.* (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).
10. Hirst, A. C., Gordon, H. B. & O'Farrell, S. P. *Geophys. Res. Lett.* **23**, 3361–3364 (1996).
11. Knutson, T. R., Delworth, T. L., Dixon, K. & Stouffer, R. J. *J. Geophys. Res.* (submitted).
12. Hegerl, G. *et al.* *Clim. Dyn.* **13**, 631–634 (1997).
13. Oppenheimer, M. *Nature* **393**, 325–332 (1998).
14. Boer, G. J., Flato, G. M. & Ramsden, D. *Clim. Dyn.* (submitted).

Antibiotic resistance

A vancomycin surprise

Michael S. Gilmore and James A. Hoch

Vancomycin is the last line of defence against organisms such as *Streptococcus pneumoniae* and *Enterococcus*, some strains of which are resistant to most other antibiotics. Whereas *S. pneumoniae* is responsible for diseases such as pneumonia, meningitis and otitis media (ear infection), *Enterococcus* bacteria cause many hospital-acquired infections. Because pneumococcal pneumonia alone results in over a million deaths each year worldwide¹, any compromise in our ability to treat these infections is an important public-health concern. In just ten years, vancomycin resistance has spread among the enterococci from one initial observation to the situation now, where 52% of clinical *E. faecium* isolates are resistant². This rapid spread seems to be due, in part, to an association between resistance to vancomycin and resistance to penicillin and ampicillin³. On page 590 of this issue, Novak and colleagues³ report a new threat to antibiotic therapy — the emergence of penicillin-resistant clinical isolates of *S. pneumoniae*, which are also tolerant to a number of other antibiotics, including vancomycin.

Tolerance and resistance are somewhat different. Whereas antibiotic-resistant microorganisms are insensitive to the antibiotic, and continue to grow in its presence, the antibiotic-tolerant strains stop growing but do not die in the presence of the antibiotic. In neither case does antibiotic therapy eliminate the infective agent, meaning that the infection can continue once therapy is curtailed. Antibiotic tolerance is particularly insidious because it cannot be detected using conventional *in vitro* tests — tolerant strains seem to be sensitive to antibiotics.

Novak *et al.*³ now show that roughly 3% of the clinical *S. pneumoniae* isolates they studied were tolerant to antibiotics. *Streptococcus pneumoniae* is the most common cause of bacterial meningitis in the United States, with the infection most prevalent among the very young and the very old⁴. Antibiotic-tolerant strains of the bacterium are most likely to cause problems at places where the body's immune response is limited, such as in the brain (meningitis) or the eyes (intraocular infection). Consistent with this proposal, the authors show that multi-

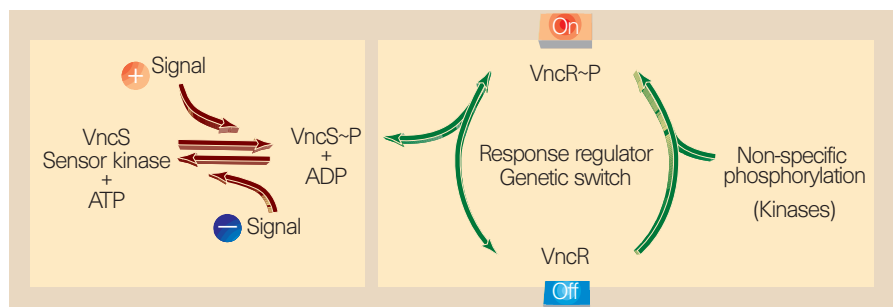


Figure 1 A two-component signal-transduction system. Environmental signals regulate the addition of a phosphoryl group (P) to the sensor kinase (VncS). This, in turn, controls whether the response regulator (VncR) is on or off. The sensor kinase can turn the response regulator off by removing its phosphoryl group in the absence of the environmental signal. It can also eliminate activation by non-specific phosphorylation of the response regulator in this way. Novak *et al.*³ have found that multiple-antibiotic-resistant bacteria have mutations in such a two-component system. When VncR is phosphorylated (active), genes that are turned on in response to antibiotics — and kill the bacteria — are switched off. Mutations in the VncS component mean that VncR cannot be switched off, with the result that the death-promoting genes remain inactive.

ple-antibiotic-tolerant pneumococci selectively persist in experimental meningitis. Moreover, in a controlled comparison, the authors found that these mutants are better able to take up DNA than normal cells. This raises the real possibility that, through DNA uptake and recombination in nature, new strains of pneumococci will continue to arise that are increasingly difficult to treat with existing drugs.

Novak and colleagues have also pinned down the genetic basis for the vancomycin tolerance — specific mutation of a two-component signal-transduction system. Two-component systems control a variety of responses in bacteria and lower eukaryotes⁵. These signal-transduction systems allow the microorganisms to sense their environment and to respond to this by adjusting gene expression. The environmental stimulus is recognized by a specific 'sensor' kinase, which, on binding the stimulating molecule, adds a phosphoryl group to itself (autophosphorylation). Each sensor kinase in the cell is paired with a specific response-regulator protein that controls the expression of a unique group of genes. Transfer of the phosphoryl group from the stimulated kinase to the response regulator activates this switch function, and the response regulator then turns gene function either on or off.

This genetic switch is active as long as the stimulating molecule is present and the sensor kinase remains phosphorylated. But if the stimulating molecule disappears, the process is reversed and the switch is turned off. The forward and reverse reactions combine to fine-tune the levels of activated response regulators present to the level of the stimulating molecule. The reverse reaction can also switch off the response regulator if it is activated, not by the sensor, but non-specifically by other kinases or by small molecules such as acetyl phosphate. Stimulating molecules may ultimately promote either phosphorylation or dephosphorylation of

the response regulator, depending on how a particular system is configured.

Novak and colleagues³ found that experimentally induced penicillin-tolerant *S. pneumoniae* mutants, which are also tolerant to structurally unrelated antibiotics such as vancomycin, aminoglycosides and quinolones, contain mutations in a two-component system (Fig. 1). The bacteria became tolerant when the authors inactivated the sensor kinase (VncS) of this system, but not when they turned off its response regulator (VncR). Consistent with this, the authors detected vancomycin-tolerant clinical *S. pneumoniae* isolates with mutations in the VncS sensor kinase. These data indicate that, in the 'on', phosphorylated mode, the VncR response regulator represses some of the genes required for antibiotic-induced death. Normally, exposure to antibiotics causes VncS to remove the phosphoryl group from VncR, returning it to the 'off' mode and allowing these genes to be expressed. But once VncR is activated (by, say, non-specific phosphorylation in the cell), it cannot then be inactivated (dephosphorylated) without functional VncS, so the antibiotic-induced genes will remain switched off.

The data also suggest that, on exposure to all of the antibiotics tested, *S. pneumoniae* produces a common signalling substance that is sensed by VncS, triggering cell death. This signalling substance could come from bacterial cell-wall precursors that cannot be incorporated into the cell wall in the presence of antibiotic. Or, it may originate from other pathways affected, directly or indirectly, by antibiotics. Whatever the mechanism, by understanding the connection between sensing the triggering molecule, and the pathway by which cell death is induced, we may be able to develop new approaches for treating increasingly common and problematic multiple-antibiotic-resistant infections⁶. For it is now clear that signal



100 YEARS AGO

The season of strawberries is at hand, but doctors are full of fads, and for the most part forbid them to the gouty. Let me put heart into those unfortunate persons to withstand a cruel medical tyranny by quoting the experiences of the great Linnæus. ... in 1750 he was attacked so severely by sciatica that he could hardly make his way home. The pain kept him awake during a whole week. He asked for opium, but a friend dissuaded it. Then his wife suggested "Won't you eat strawberries?" It was the season for them. Linnæus, in the spirit of an experimental philosopher, replied, "tentabo — I will make the trial." He did so, and quickly fell into a sweet sleep that lasted two hours, and when he awoke the pain had sensibly diminished. He asked whether there were any strawberries left: there were some, and he eat them all. Then he slept till morning. ... What lucrative schemes are suggested by this narrative. Why should gouty persons drink nasty waters, at stuffy foreign Spas, when strawberry gardens abound in England? Let enthusiastic young doctors throw heart and soul into the new system. From *Nature* 8 June 1899.

50 YEARS AGO

In November 1948, we found at Swartkrans while excavating a new site in co-operation with the California University Expedition a lower jaw with very large molar teeth. I called it *Paranthropus crassidens*. Not only did we find the imperfect jaw of a young male with three upper teeth, but also some months later the snout of what appeared to be a female. The upper incisors and canine are typically human. Last month, while I was in America, my assistant, Mr. John T. Robinson, discovered an almost complete huge jaw with most of the teeth. ... The jaw is really enormous — very considerably larger than that of Heidelberg man. The molar teeth are very large, but the canines and incisors remarkably small. Another remarkable feature about the jaw is that it has a rudiment of a chin. The ascending ramus is also very high and large. Mr. Robinson has also got a part of the face with the nearly complete palate. This shows that the face is almost orthognathous. We must, I think, conclude that the brain was very large. From *Nature* 11 June 1949.

transduction is essential both in bacterial virulence and in antibiotic susceptibility⁷. □

Michael S. Gilmore is in the Departments of Ophthalmology and Microbiology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190, USA. James A. Hoch is in the Division of Cellular Biology, Department of Molecular and Experimental Medicine, Scripps Research Institute, La Jolla, California 92037, USA. e-mail: hoch@scripps.edu

1. WHO Global Programme for Vaccines and Immunization, Vaccine Research and Development Rep. Tech. Rev. Group Mtg

- 10 (WHO, Geneva, 1998).
2. Huycke, M. M., Sahn, D. F. & Gilmore, M. S. *Emerg. Infect. Dis.* **4**, 239–249 (1998).
3. Novak, R., Henriques, B., Charpentier, E., Normark, S. & Tuomanen, E. *Nature* **399**, 590–593 (1999).
4. Centers for Disease Prevention and Control *Prevention of Pneumococcal Disease: Recommendations of the Advisory Committee on Immunization Practices* 1–24 (MMWR 46 (RR-08), 1997).
5. Swanson, R. V., Alex, L. A. & Simon, M. I. *Trends Biochem. Sci.* **19**, 485–490 (1994).
6. Barrett, J. F. & Hoch, J. A. *Antimicrob. Agents Chemother.* **42**, 1529–1536 (1998).
7. Dziejman, M. & Mekalanos, J. J. in *Two-Component Signal Transduction* (eds Hoch, J. A. & Silhavy, T. J.) 305–317 (ASM Press, Washington DC, 1995).

Superconductivity

Cohabitation in the cuprates

Arthur P. Ramirez

Like many a marriage between opposites, the confluence of superconductivity and magnetism in the layered cuprates has produced interesting offspring. Superconductivity and magnetism are usually considered mutually exclusive because the superconducting state consists of pairs of electrons with anti-aligned spins. A large enough magnetic field can align the spins and, in so doing, destroy the superconducting state. In a forthcoming paper¹ in *Physical Review B*, a group led by Bob Birgeneau writes a new chapter in the history of superconductors by showing that, in an exquisite lanthanum cuprate crystal, the temperature where superconductivity sets in, T_c , is exactly the same as where a static internal magnetic field develops (Fig. 1). Such behaviour renews the question of whether magnetism is beneficial or harmful to superconductivity in these materials.

When a magnetic field is applied to a common (type II) superconductor, individual field lines thread through the sample, and around each line is a small region where superconductivity vanishes. Therefore, for values of the applied field lower than that needed to destroy all the electron pairs, normal metallic regions coexist with superconducting regions. In many superconductors, these internal fields can be much larger than those produced in common magnets. So it is possible that, under certain conditions, magnetic moments can coexist with superconductivity — such a coexistence might be possible in hybrid compounds possessing two sets of atoms, one set contributing conduction electrons and the other magnetic moments. In such a hybrid, the internal field produced by the magnetic ions threads through the sample like that of an applied field, coexisting with the superconducting electrons. Such materials do exist. For instance, in the rare-earth nickel–boron carbides an antiferromagnetic state — where the magnetic moments pointing in one direction are balanced by others pointing in the opposite direction — among the

rare-earth atoms coexists with superconducting electrons originating from the other atoms. This spatial dichotomy between the atoms contributing magnetism and the atoms contributing superconducting electrons is analogous to sodium and chlorine coexisting in distinct but periodic regions of salt crystals.

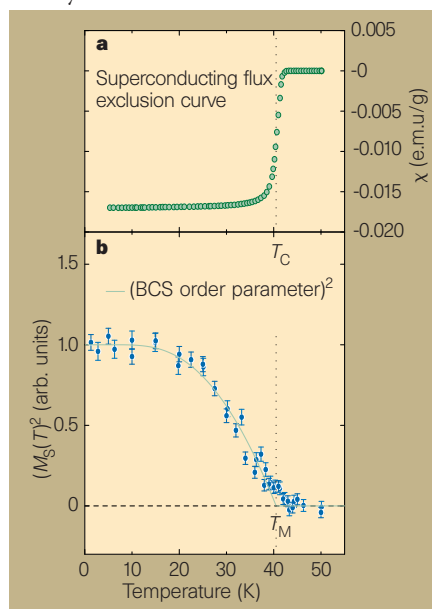


Figure 1 The confluence of superconductivity and magnetism in a high- T_c cuprate superconductor. a, Bulk magnetic susceptibility, χ , showing the diamagnetic shift associated with superconductivity, in an oxygen-intercalated sample of lanthanum cuprate. The transition temperature, T_c , is the dotted line. b, The data points represent the strength of internal magnetism, which is proportional to the square of the magnetic order parameter M_s . The associated magnetic field starts to appear at the same temperature as the superconductivity ($T_m = T_c = 40.5$ K). To stress the relationship between the two normally incompatible states, a solid line describing a common type of superconductivity, known as Bardeen–Cooper–Schrieffer (BCS) pairing, is also shown.

The challenge posed by superconductivity in high- T_c cuprates arises from the dual role of the copper atoms. In the parent compound lanthanum cuprate, which is an insulator, the copper–copper interaction leads to antiferromagnetism. When this material is doped with 2% electron vacancies (holes) per copper atom — for example, by substituting divalent strontium for trivalent lanthanum, or by introducing oxygen into interstitial sites of the crystal — antiferromagnetism is destroyed². For higher concentrations of holes, superconductivity occurs. The dopant atoms are not replacing magnetic ones, so in this system the same atom, copper, is responsible for both superconductivity and magnetism. The question that has motivated much research on these materials is what happens to the magnetism when the material is converted to a superconductor by doping? The clearest answers to this question have come from neutron scattering, because neutrons are unique in their ability to probe both the energy and wavelength dependence of magnetic excitations.

The Birgeneau group¹ has found, using neutron scattering, that antiferromagnetism actually coexists with superconductivity even though both phenomena originate from the same atom. The key advance in this work was the preparation of crystals in which the dopant atoms themselves, oxygen in this case, are arranged periodically. The periodic arrangement is made possible by oxygen's high mobility. Still, achieving this degree of order was not easy and the crystal yielding the data in Fig. 1 was oxygenated over several weeks. This tedious process allowed the excess oxygen atoms to find a pattern in which every fourth copper layer has excess oxygen interwoven into the structure — a stunning feat of materials engineering.

There is an interesting twist to the data in Fig. 1. The wavelength of the antiferromagnetism, whose strength is shown in Fig. 1b, is not a rational fraction of the atomic lattice spacing. This 'incommensurate' magnetism results from the frustration of the magnetic moments in trying to achieve a low-energy configuration within the constraints imposed by the crystal lattice. This incommensurability had been known to exist in the doped cuprates^{3–6} and also to be related to superconductivity⁷. However, previous neutron-scattering measurements had been done on samples exhibiting a T_c much lower than the largest observed value. Because a low T_c can be due to crystal imperfections, it was not clear whether the simultaneous occurrence of static magnetism and superconductivity was an intrinsic effect or not. The new measurements involve a T_c among the highest seen for doped lanthanum cuprate, showing that the effect occurs even under optimum conditions for superconductivity.

The value of T_c increases with increasing

incommensurability in the layered cuprates, which argues for competition between magnetism and superconductivity. In the present work, it seems possible that oxygen intercalation has produced two types of copper atoms, those close to the dopants and those far away, bringing us back to a conventional view of the need for a spatial dichotomy to accommodate both magnetism and superconductivity. Another possible explanation involves holes spatially segregated within each plane, forming charge stripes or spatially distinct regions of different electronic character⁸. On the other hand, many other measurements support a view in which the relationship is not adversarial, but where magnetic fluctuations actually induce superconductivity in the cuprates⁹. Such theories hark back to work done in the 1980s on 'heavy-fermion' superconductivity¹⁰. Like the cuprates, heavy-fermion systems exhibit a delicate interplay of magnetism and superconductivity, although with $T_c \sim 1$ K. Also, heavy-fermion phenomena are associated with the same atomic species, uranium, and sometimes occur in the same sample¹¹. This

suggests parallels between these two very different types of metals, copper oxides and uranium intermetallics. The experiments of Birgeneau and colleagues¹ show that an equitable living arrangement can be reached between two unlikely partners and present a clear challenge for theories of high- T_c superconductivity. □

Arthur P. Ramirez is at the Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA.

e-mail: apr@bell-labs.com

1. Lee, Y. S. *et al.* *Phys. Rev. B* (in the press); <http://xxx.lanl.gov/abs/cond-mat/9902157>
2. Kastner, M. A., Birgeneau, R. J., Shirane, G. & Endoh, Y. *Rev. Mod. Phys.* **70**, 897–928 (1998).
3. Yoshizawa, H., Mitsuda, S., Kitazawa, H. & Katsumata, K. *J. Phys. Soc. Jpn* **57**, 3686–3689 (1988).
4. Birgeneau, R. J. *et al.* *Phys. Rev. B* **39**, 2868–2871 (1989).
5. Dai, P., Mook, H. A. & Dogan, F. *Phys. Rev. Lett.* **80**, 1738–1741 (1998).
6. Mook, H. A. *et al.* *Nature* **395**, 580–582 (1998).
7. Kimura, H. *et al.* *Phys. Rev. B* **59**, 6517–6523 (1999).
8. Kivelson, S. A., Fradkin, E. & Emery, V. J. *Nature* **393**, 550–553 (1998).
9. Pines, D. Z. *Phys. B* **103**, 129–135 (1997).
10. Schmitt-Rink, S., Miyake, K. & Varma, C. M. *Phys. Rev. Lett.* **57**, 2575–2578 (1986).
11. Aepli, G. *et al.* *Phys. Rev. Lett.* **63**, 676–679 (1989).



CORBIS/STEVE KAUFMAN

Figure 1 Hammering home the point — the acorn woodpecker (*Melanerpes formicivorus*). Koenig and Haydock¹ have found that the population stability of these woodpeckers is limited not by overall density of the oak trees from which they gather their acorns, but by the diversity of oak species.

Ecology

Woodpecker population drills

Peter D. Moore

The proposal that species diversity in an ecosystem makes it more stable — that is, more resistant to perturbation — is intuitively attractive to ecologists. But this idea is not always supported by either natural or model systems. One theoretical argument in favour of the proposal is that a predator faced with declining levels of one prey species could switch to alternatives. This strategy would reduce the pressure on the original prey population and allow it to recover. A range of prey options would also mean that the population of predators should not be subject to violent fluctuation if one species of prey periodically became scarce. According to a report by Walter Koenig and Joseph Haydock¹ in the *Journal of Biogeography*, this principle seems to operate in populations of the acorn woodpecker (*Melanerpes formicivorus*; Fig. 1) on the Pacific coast of California.

Acorn woodpeckers are blessed with a number of biological peculiarities. They breed communally in groups of 12–15, yet only two or three pairs, which share mates between them, are involved in the breeding². The remainder of the group act as helpers in gathering food, defending the territory and caring for the young. The young are initially fed on insects, but adults prefer acorns gathered from a range of available oak species. The woodpeckers store their acorns in 'granaries' — consisting of an old tree or often a telegraph pole — by drilling holes and then

hammering in the acorns.

Acorn woodpeckers are abundant along the coastal fringe of California, although they are not found in the deserts of eastern California. Their range extends north into Oregon, and even to the offshore islands of Santa Cruz and Santa Catalina³. They also live, at lower densities, in wooded regions of Arizona, New Mexico, Texas and south into Mexico. Coastal California is rich in oak trees, including evergreen species (such as *Quercus agrifolia* and *Q. chrysolepis*) and deciduous species (including *Q. lobata* and *Q. douglasii*). All produce acorns that are enthusiastically gathered by the acorn woodpecker.

The availability of diverse sources of food for such an acorn-dependent bird has led ecologists and biogeographers to ask whether the population density of these woodpeckers is controlled by the abundance of oaks in general, or by the diversity of species. Work in 1974 by Bock and Bock⁴ showed that overall oak abundance and woodpecker density were positively correlated. However, the limits of acorn woodpecker distribution coincided not with the total number of oak trees, but with the species diversity. Koenig and Haydock¹ have now re-examined this relationship with the aid of 30 years' worth of data from the Audubon Society's Christmas bird counts, and using information about the pattern of fruiting in oak trees.

The authors first considered the possibility that a richer range of species might provide a more reliable supply of acorns. Oak trees vary in their acorn production — some species require only a year for acorns to mature whereas others need two. So, a more diverse assemblage of oaks could provide a more reliable resource in the long term. Koenig and Haydock found that this was generally the case, the effect being particularly apparent when the diversity of oak trees dropped to a single species. The variability in acorn production for a rich assemblage gradually rose as the number of species declined. But the difference between two species and a single species, in terms of the reliability of acorn production, was statistically significant.

Census data for woodpeckers from both coastal and inland regions confirmed the earlier findings of Bock and Bock — that woodpecker populations cannot be maintained if only one species of oak is present. This is clearly explained by problems of reliable food supply. For two species of oak, the likelihood of crop failure is only about 12% (effectively once every two woodpecker generations). With only one species, however, the probability of failure rises to 24% (once every generation), and this precludes the long-term survival of woodpecker populations.

Further analysis of the coastal Californian data showed that the overall density of acorn woodpeckers is closely related to the abundance of oaks, irrespective of species. But Koenig and Haydock found that the annual variability of woodpecker populations correlates more closely with the number of species. These findings confirm the

intuitive expectation that the population density is determined by overall food resources, whereas fluctuations within the population are controlled by diversity of the food supply (that is, the number of oak species). In Mexico and inland areas of the United States, where the density of acorn woodpeckers is much lower, abundance of the birds again seems to be related to the overall density of oak trees. However, the effect of species diversity is not apparent in this region. Some other factor is probably important in regulating woodpecker populations in this part of their range.

The most influential change in a woodland, as far as the acorn woodpecker is concerned, is the increase from one species of oak to two. So, would adding further species be redundant to efficient functioning of the system? The addition of more species

enhances the fail-safe mechanism that ensures long-term stability for the woodpecker population. In other words, although additional species may not be 'rivets', in the sense that each one adds equally to overt efficiency of the system⁵, neither can they be dismissed as redundant. They are better regarded as additional back-ups — insurance policies against the catastrophe that would result from failure within the system. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

e-mail: peter.moore@kcl.ac.uk

1. Koenig, W. D. & Haydock, J. J. *Biogeogr.* **26**, 159–165 (1999).
2. Schoenherr, A. *A Natural History of California* (Univ. California Press, Berkeley, 1992).
3. Small, A. *California Birds: Their Status and Distribution* (Ibis, Vista, California, 1994).
4. Bock, C. E. & Bock, J. H. *Am. Nat.* **108**, 694–698 (1974).
5. Lawton, J. H. *Oikos* **71**, 367–374 (1994).

Condensed-matter physics

Phases of resistance

Peter Littlewood

A familiar and convenient classification of materials is between metal and insulator, and there are even some rare compounds where one can transform into the other at modest pressures or temperatures. The class of magnetic oxides described by Uehara *et al.*¹ on page 560 of this issue seems to be ambivalent; the material has co-existing phases of metal and insulator with identical chemical composition, and offers a new route to 'colossal' magnetoresistance. This is the effect seen when a small magnetic field gives rise to a large change in electrical resistance. It might turn out to be technologically useful, for example allowing an increase in the density of data that can be stored in magnetic media.

Manganese perovskites based on LaMnO_3 have been attracting considerable scientific and technological interest lately on account of their remarkable electronic and magnetic properties. Many of these materials are ferromagnetic metals (similar to iron), which on heating lose their magnetic order, and this is accompanied by a huge increase in electrical resistance. Applying a small magnetic field to this poorly conducting material restores some ferromagnetic alignment causing the resistance to drop substantially, a phenomenon that has been dubbed 'colossal magnetoresistance' (CMR), because it is much larger than that observed in standard magnetic materials². So CMR arises from competition between a metal and an insulator, but the character of the insulating state has been unclear until now.

There are many ways to manipulate this metal–insulator transition. One route is to make partial chemical substitution of the

trivalent La^{3+} . Replacing La with a divalent ion (such as Sr^{2+} or Ca^{2+}) will vary the proportion of Mn^{3+} and Mn^{4+} (that is, change the nominal Mn valence); substituting with a smaller trivalent rare-earth atom (such as Pr^{3+}) will change the lattice constant, producing a kind of chemically induced pressure. Changing the Mn valence adds potentially mobile charge carriers, whereas smaller cations tend to induce insulating behaviour.

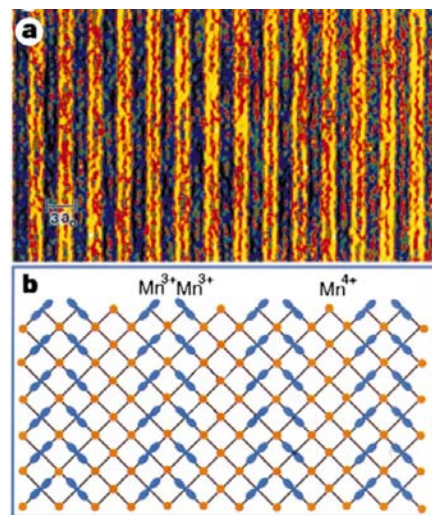


Figure 1 Charge-ordered stripes seen in $\text{La}_{0.33}\text{Ca}_{0.67}\text{MnO}_3$ (adapted from ref. 3). Charge ordering is a feature of transition-metal oxides, such as the magnetoresistive compounds studied by Uehara *et al.*¹. a, The high-resolution image made by electron microscopy shows pairing of stripes with a period of three times the fundamental lattice constant. b, This is interpreted as a periodic ordering of Mn^{3+} and Mn^{4+} ions, with the occupied *d*-orbitals shown in blue and orange respectively.

When the nominal Mn valence is intermediate between $3+$ and $4+$, at least three different states of the solid can be seen: a ferromagnetic metal, a charge-ordered insulator, and at higher temperatures a poorly conducting paramagnet.

In the ferromagnetic state the electrons are mobile and the charge density on each Mn site is identical; the mobile electrons promote ferromagnetism by aligning the core spins on the Mn ions in a process called double exchange. Alternatively, the charge segregates preferentially between different sites with nominally Mn^{3+} or Mn^{4+} valence, producing a charge-ordered state³, which is also an insulator (Fig. 1). Such charge-ordered structures are not only seen in manganese oxides, but also in cuprates⁴ and nickelates⁵; charge stripes are a common (but recently appreciated) phenomenon of transition-metal oxides. At high enough temperatures this ordered array can 'melt' and become dynamic, while retaining local charge fluctuations. This liquid-like paramagnetic state is sensitive to magnetic fields and this is when CMR becomes most pronounced⁶.

The stability of one phase over another is a matter of chemistry and can be tuned. Uehara *et al.*¹ study a series of compounds, where the La is partially substituted by the smaller Pr, but the nominal Mn valence is fixed. The Pr-rich compounds have stable charge-ordered phases and the La-rich compounds are ferromagnetic⁷, so one would expect to see a sudden insulator-to-metal transition as the Pr:La ratio varies. They find instead a broad range over which the two phases coexist. Electron microscopy shows that both metal and insulator are present in the same crystal, segregated on length scales of hundreds of nanometres to micrometres. Other workers⁸ have now imaged similar structures in a different compound, $(\text{La,Ca})\text{MnO}_3$. As the La:Pr ratio is varied in the $(\text{La,Pr,Ca})\text{MnO}_3$ system, only the proportion of metal to insulator changes. Here, CMR is produced by the coexistence of two phases, not by the homogeneous transformation of one to the other, as observed with changing temperature in conventional manganites.

Why should two-phase coexistence be surprising? It is commonly produced by chemical segregation, but Uehara *et al.* show that this is minimal in their samples. Another old idea is of electronic charge segregation into magnetic clusters; but micrometre-sized domains are too large for large-scale charge separation with its huge penalty in electrostatic energy. Although the phases coexist over wide variations in composition, it is known that the charge-ordered phase can be transformed to the ferromagnetic state by modest magnetic fields, and even by substitution of ^{16}O with ^{18}O . This implies that the ground-state energies of the two phases remain very close, despite large compositional changes. Coexistence of metal and

insulator in this way is unprecedented.

A familiar example of coexistence of different phases — the liquid–gas transition of water — might be a useful analogy here. Liquid water and water vapour have different natural densities that depend on pressure. If one places some water in a sealed and evacuated container at room temperature, liquid and vapour will come to equilibrium at the same pressure. If the volume or the temperature is changed, the proportions of the two phases will adjust, despite the fact that the liquid phase is the thermodynamic ground state at fixed pressure. Constraining the system in a fixed volume enforces coexistence.

How might this relate to CMR? The ferromagnetic and the charge-ordered phases have a large strain mismatch. So if part of a single ferromagnetic crystallite nucleates into the charge-ordered phase, that domain is under a huge stress from the surrounding crystal that discourages further growth. Domains in different parts of the crystal will form with strain fields in random orientations, so the inhomogeneous stress is not easily released. The crystal will then evolve to a glassy pattern of charge-ordered domains with different orientations, coexisting with untransformed metallic domains. Such situations are not uncommon in ferroelectric and ferroelastic oxides⁹, where of course the competing phases are usually non-metallic. The parallel between disordered and 'relaxor' ferroelectrics has recently been solidified by measurements of hysteresis and slow dynamics on several mixed-phase manganites^{8,10}.

Whatever the origin, there are substantial consequences. Measurement of electrical resistance reveals a metal or an insulator, depending on whether or not the metallic grains form a percolating network through the sample. External perturbations, such as magnetic field, light, X-rays or stress, change the relative proportions of the ferromagnetic and charge-ordered phases. In a related observation, Fiebig *et al.*¹¹ have shown that well-conducting paths are created simply by shining light on insulating $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. And the enormous effects of imposing lattice strain on CMR have long been a puzzle, dating back to some of the earliest results on thin films¹². Moreover, this new phenomenon is not unique to a single compound, having now been seen in several other materials^{8,10,13}.

The richness of self-organized patterns in condensed-matter systems always exceeds our ability to predict them. Such interpenetrating structures of metal and insulator seem almost designed for technological purposes, and there have already been studies on deliberately patterned and fabricated junctions with a view to device applications. Nature is delivering such structures for free.

Peter Littlewood is at the Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK.

e-mail: peter.littlewood@phy.cam.ac.uk

1. Uehara, M., Mori, S., Chen, C. H. & Cheong, S.-W. *Nature* **399**, 560–563 (1999).
2. Ramirez, A. P. *J. Phys.: Condens. Matt.* **9**, 8171–8199 (1997).
3. Mori, S., Chen, C. H. & Cheong, S. W. *Nature* **392**, 473–476 (1998).
4. Tranquada, J. M. *et al.* *Nature* **375**, 561–563 (1995).
5. Chen, C. H. *et al.* *Phys. Rev. Lett.* **71**, 2461–2464 (1993).
6. Millis, A. J. *Nature* **392**, 147–150 (1998).
7. Tomioka, Y. *et al.* *Phys. Rev. B* **53**, R1689–R1692 (1996).

8. Kimura, T. *et al.* (Joint Research Centre for Atom Technology, Tsukuba, Japan, preprint).
9. Strukov, B. A. & Levanyuk, A. P. *Ferroelectric Phenomena in Crystals* 193–224 (Springer, New York, 1998).
10. Barnabe, A. *et al.* *Appl. Phys. Lett.* **71**, 3907–3909 (1997).
11. Fiebig, M., Miyano, K., Tomioka, Y. & Tokura, Y. *Science* **280**, 1925–1928 (1998).
12. Jin, S. *et al.* *Science* **264**, 413–415 (1994).
13. Fäth, M. *et al.* (Kamerlingh Onnes Lab., Leiden Univ., preprint).

Immunology

Looking out for memory T cells

Benedita Rocha

After infection or vaccination, many organisms can 'remember' the identity of the pathogen (antigen) responsible, allowing them to respond with enhanced vigour and protecting them from re-infection, sometimes throughout life¹. This phenomenon, known as immunological memory, depends on antigen-specific lymphocytes — the long-lived 'memory' T cells that are generated during the first encounter with a particular antigen. If we could identify T cells with memory potential (that is, the capacity for long-term survival), this would be extremely useful for assessing vaccinating protocols, and, more generally, for investigating immune responses. Jacob and Baltimore², reporting on page 593 of this issue, claim that this can now be done.

During the first response to a pathogen, many antigen-specific T cells are generated but only a minority survive. The kinetics of such responses can be divided into three distinct phases¹. The first is expansion, in which antigen-specific lymphocytes are activated to divide. The number of lymphocytes rapidly becomes enormous, and the lymphoid organs (such as the spleen and lymph nodes) enlarge to accommodate them. The activated T cells then begin to take on 'effector' functions; they secrete cytokines (peptides that signal between lymphoid and non-lymphoid cells), and kill infected target cells. The second phase is contraction. Division and the effector functions of the activated T cells subside as the antigen — the driving force behind these phenomena — is gradually eliminated. Over 95% of the antigen-specific T cells then die, and the lymphoid organs return to their normal size. Finally comes the memory phase, in which those T cells that have been spared by the contraction phase survive for long periods, forming a stable pool of 'memory' cells.

In such immune responses, how can we identify those T cells that will survive and generate a stable memory pool? To address this question, Jacob and Baltimore² used a genetic recombination system³ mediated by the Cre recombinase to 'tag' activated T cells with a reporter gene, human placental alkaline phosphatase (PLAP). The Cre recombinase is carried on one transgenic construct, and is activated only in activated T cells.

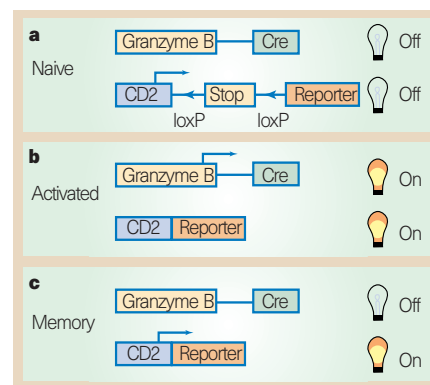


Figure 1 Strategy used by Jacob and Baltimore² to label activated T cells *in vivo*. Mice have two transgene constructs. In the first, the granzyme B promoter drives expression of the Cre recombinase. Granzyme B is an enzyme used by activated T cells to kill infected cells, and its promoter is active only in activated T cells. The second, reporter transgene contains a T-cell promoter (CD2), a stop codon flanked by substrates for Cre-mediated recombination (loxP), and a reporter gene (PLAP). a, In naive T cells the granzyme B promoter is silent so cells do not transcribe the Cre recombinase. The CD2 promoter is active but the stop codon prevents expression of PLAP. b, In activated T cells, the granzyme B promoter is active so cells express the Cre recombinase. This enzyme removes the stop codon from the reporter transgene so PLAP is expressed. Once this stop codon has been removed, PLAP is permanently expressed. c, Memory cells contain PLAP, but they no longer express the Cre recombinase.

Once Cre is activated in these cells it removes a stop codon from a second transgenic construct, allowing PLAP to be expressed. In principle, this method should identify naive, or unactivated, T cells (in which neither Cre nor PLAP is expressed; Fig. 1a); activated T cells (which should produce both Cre and PLAP; Fig. 1b); and memory T cells (which will contain the PLAP marker, but no longer express the Cre recombinase; Fig. 1c).

In theory, then, all activated T cells should become indelibly marked with PLAP. Yet Jacob and Baltimore found that, in fact, only a minority of the activated T cells in transgenic mice expressed this gene. We know that about half of the T lymphocytes

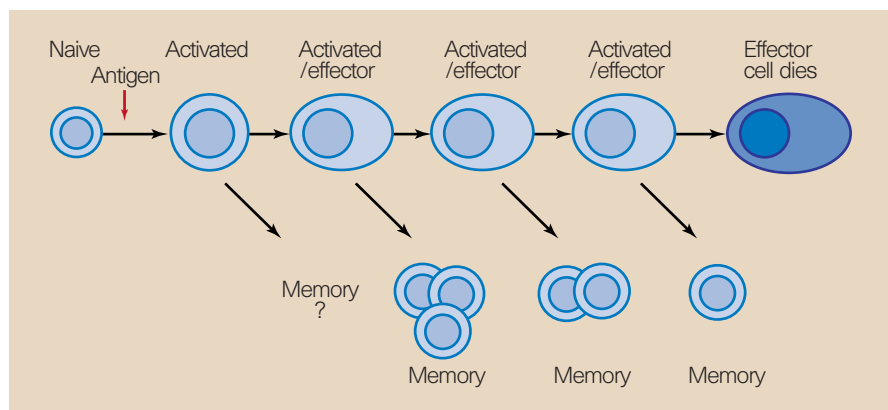


Figure 2 Model for the development of memory T cells. When activated by a specific antigen, T cells divide and differentiate, and start to carry out effector functions. Such activated/effector cells can revert to memory cells. But successive cell divisions reduce the chances that a T cell will develop into a memory T cell, and increase the chances that it will die.

in peripheral areas such as the spleen and lymph nodes express memory markers⁴, yet the authors found no PLAP-containing T cells in these areas. They found PLAP only in the Peyer's patches (collections of lymphocytes in the wall of the small intestine). To investigate possible differences between the T cells that expressed PLAP and those that didn't, Jacob and Baltimore studied induction of PLAP expression in a well-characterized model — *in vivo* infection with lymphocytic choriomeningitis virus (LCMV).

When antigen determinants (specific peptide sequences that the immune system recognizes) are known, antigen-specific T cells can be identified *in vivo*. These peptides can be complexed to tetrameric molecules of the major histocompatibility complex, and labelled with fluorochromes. Antigen-specific T cells then bind to these labelled complexes. This method allows the kinetics of antigen-specific T cells to be followed after immunization, and others⁵ have used it to characterize the immune response after infection with LCMV. In the naive animal, LCMV-specific T cells are so rare they cannot be identified. Viral infection is followed by massive proliferation of these T cells — at the peak of the response, as many as 70–90% of the total CD8-positive T cells are specific to LCMV. But once the virus has been eliminated, most of these antigen-specific T cells do not survive the contraction phase, with only about 5% becoming memory T cells⁵.

When following PLAP expression in mice infected with LCMV, Jacob and Baltimore² found that only about 10% of the LCMV-specific T cells become PLAP-positive during the acute phase of the response. So what is the difference between the LCMV-specific T cells that contain PLAP and those that don't? Both populations seemed to contain effector T cells, as both could kill target cells, but they differed in their long-term survival. Whereas at least some of the PLAP-containing cells could proliferate *in vitro* in response to antigen,

those that did not contain PLAP simply died. Most importantly, PLAP expression was associated with the ability to become memory T cells — half of the cells that contained PLAP survived *in vivo* in this form. Although some memory cells lacking PLAP could have formed, most of the T cells that survived the contraction phase to become memory T cells probably contained PLAP.

Why did only a fraction of the activated T cells express the transgenic Cre recombinase and produce PLAP? We don't know, but studies of the physiological differences between those T cells that contain PLAP and those that do not will stimulate a new field of research into the differences in T-cell behaviour after they have been stimulated with antigen *in vivo*. Interestingly, the T cells that subsequently became memory cells proliferated relatively weakly during the expansion phase of the immune response. This observation supports a model of T-cell differentiation in which increasing cell stimulation and division are associated with progress to terminal differentiation and a reduction in the memory potential (Fig. 2). In other words, cells that divide many times are more likely to die than to survive as memory cells.

Jacob and Baltimore's results indicate that we should now be able to identify T cells with 'memory' potential from among the bulk of antigen-reactive T cells. Effective vaccination depends on the generation of memory cells — so if cells with memory potential express PLAP, this expression may be used to identify efficient immunization protocols. And, if confirmed in other infection models, the approach could be invaluable for assessing immune responses. □

Benedita Rocha is at the Institut Necker, INSERM U 345, 156 rue de Vaugirard, 75015 Paris, France. e-mail: rocha@necker.fr

1. Ahmed, R. & Gray, D. *Science* **272**, 54–60 (1996).
2. Jacob, J. & Baltimore, D. *Nature* **399**, 593–597 (1999).
3. Lasko, M. et al. *Proc. Natl Acad. Sci. USA* **89**, 387–419 (1995).
4. Tanchot, C. & Rocha, B. *Immunol. Today* **19**, 575–579 (1998).
5. Murali, K. K. et al. *Immunity* **8**, 177–187 (1998).

Daedalus

Drawn from life

Optical and electronic communication uses vast amounts of bandwidth, essentially for pictures. Clever coding can compress this information, but must still retain enough to reconstruct the whole visual extravaganza at the far end. Yet in many cases pictures convey very little of value, even assuming they are necessary at all. Indeed, DREADCO's information scientists are now writing a special NDP ('No Damned Pictures') program for Internet users. It will reject the slow-downloading visual nonsense, while accepting elements of text that actually have something to say.

But even Daedalus regards this as bit extreme. He is looking for a middle way. He points out that a good cartoonist can capture the essence of a scene with a few well-chosen lines. Why is a line drawing so direct and easy to interpret? Our visual system must work largely on edge detectors; yet this cannot be the whole story. A reversed white-on-black line drawing is not nearly so easy to understand. But clearly a vast amount of bandwidth could be saved if the full bit-map of an image could be reduced to an equivalent line drawing.

So a panel of DREADCO cartoonists is making sketches of a wide variety of photographs, while psychologists study their choice of lines, and programmers strive to embody these choices in software. Their pilot program merely looks for continuous boundaries defining a sharp change in visual contrast. It renders these as black lines on a white ground, and ignores everything else. The results are promising, but imperfect. Block colours, added by an electronic 'painting by numbers' routine, seem to help a bit. An effective 'Videosketch' program will need much greater insight and subtlety; but the thing seems possible.

When perfected, Videosketch will transform business and scientific communication. It will transmit useful material, such as text, drawings, and graphs, almost unchanged. Rococo visual detail, however, will be drastically pruned. Photographs will become sketches, video clips will become cartoon films, video-conferencing will acquire an intriguing surreality. Even entertainment may benefit. Daedalus will be interested to see if video on demand, that great white hope of the cable industry, takes off in cartoon form. 'Adult entertainment' might even gain from the stark, compelling simplicity of its Videosketch representation.

David Jones

Hiding messages in DNA microdots

The microdot is a means of concealing messages (steganography)¹ that was developed by Professor Zapp and used by German spies in the Second World War to transmit secret information². A microdot ("the enemy's masterpiece of espionage"²) was a greatly reduced photograph of a type-written page that was pasted over a full stop in an innocuous letter². We have taken the microdot a step further and developed a DNA-based, doubly steganographic technique for sending secret messages. A DNA-encoded message is first camouflaged within the enormous complexity of human genomic DNA and then further concealed by confining this sample to a microdot.

A prototypical 'secret message' DNA strand contains an encoded message flanked by polymerase chain reaction (PCR) primer sequences (Fig. 1a). Encryption is not of primary importance in steganography, so we can use a simple substitution cipher¹ to encode characters in DNA triplets (Fig. 1b). Because the human genome contains about 3×10^9 nucleotide pairs, fragmented and denatured human DNA provides a very complex background for concealing secret-message DNA. For example, a secret message 100 nucleotides long added to treated human DNA at one copy per haploid genome would be hidden in a roughly three-million-fold excess of physically similar DNA strands. Confining such a sample to a microdot might then allow even the medium containing the message to be concealed from an adversary. However, the intended recipient, knowing both the secret-message DNA PCR primer sequences and the encryption key, could readily amplify the DNA and then read and decode the message.

Even if an adversary somehow detected such a microdot, it would still prove extremely difficult to read the message without knowing the specific primer sequences. For example, if 20-base random primers were used to amplify the DNA, separate amplifications with more than 10^{20} different primer pairs would be required, even allowing three mismatches per primer, followed by analysis of any PCR products obtained. Similar considerations apply to attempts to shotgun-clone the DNA sample and analyse the resultant clones. So even if the same primer pair were used on several occasions, an enemy trying to detect the primer sequences would face an extremely difficult experimental barrier. Further mathematical and biochemical analysis would therefore be expected to prove that the primer pairs used in this technique are not analogous to a classic, single-use, cryptographical "one-time pad"¹.

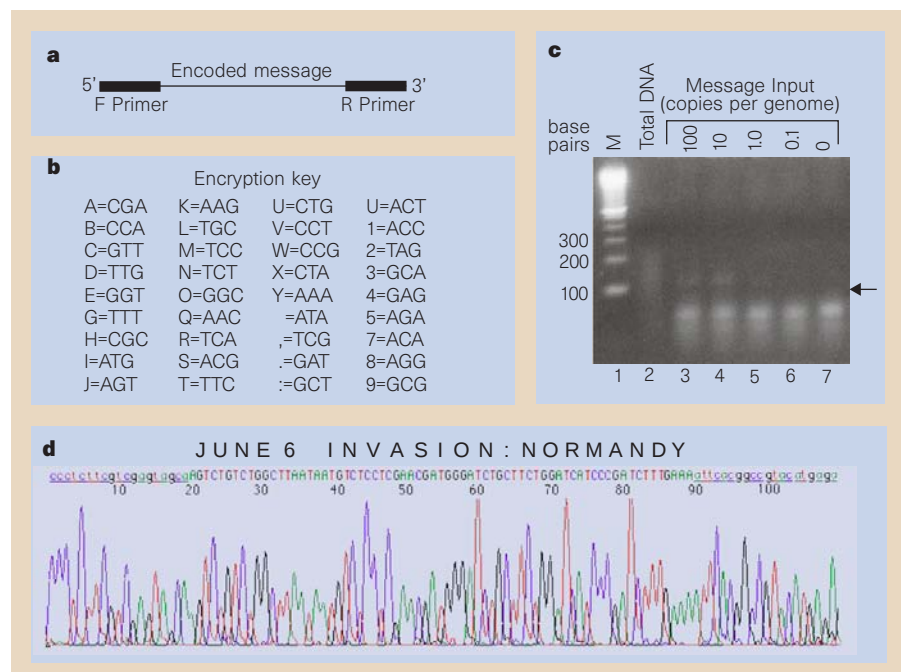


Figure 1 Genomic steganography. **a**, Structure of a prototypical secret-message DNA strand. F, forward; R, reverse. **b**, Key used to encode a message in DNA. **c**, Gel analysis of products obtained by PCR amplification with specific primers of microdots containing secret-message DNA strands hidden in a background of sonicated, denatured human genomic DNA. Message input in copies per human haploid genome is indicated, where 1.0 corresponds to 0.41 femtograms of secret-message DNA in 11 nanograms of human DNA. Lane 2 contains a message input of 100 (20-fold more total DNA than the microdots) and was not PCR amplified. M, 100-base-pair size markers. The gel was stained with ethidium bromide. The arrow indicates the PCR product seen in some lanes, below which primer-dimer bands can be seen. **d**, Sequence of the cloned product of PCR amplification, and the result of using the encryption key to decode the message. The DNA sequence determined for the encoded message is shown; the flanking primer sequences are in lower case. For details of the experimental methodology, see Supplementary Information.

Attempts by an adversary to use a subtraction technique to detect the secret-message DNA concealed within human DNA could be blocked by using a random mixture of genomic DNAs from different organisms as background. The intended recipient could still use the same procedures to amplify and read the secret-message DNA, even if ignorant of the random mixture composition, and even if the primers artefactually amplified a limited number of genomic sequences, because the encryption key would reveal which PCR product encodes a sensible message. This technique would also allow a single or duplicate microdots to be used to send individual secret messages to each of several intended recipients, each of whom would use a unique set of primers to amplify only his or her intended message.

To investigate the feasibility of this scheme, we synthesized a secret-message DNA oligodeoxynucleotide containing an encoded message 69 nucleotides long flanked by forward and reverse PCR primers, each 20 nucleotides long. We prepared concealing DNA that is physically

similar to the secret-message DNA by sonicating human DNA to roughly 50 to 150 nucleotide pairs (average size) and denaturing it. We pipetted 6 μ l of each solution containing 225 ng of treated human DNA, plus various amounts of added secret-message DNA, over a 16-point full stop printed on filter paper; it finally occupied an area about 20 times the size of the full stop. Excision of the printed full stops, each containing about 10 ng of DNA and with a cross-section that was about 75% larger than a full stop on this page, yielded DNA microdots.

Primers designed to amplify the secret-message DNA were used to perform PCR directly on DNA microdots, without prior DNA solubilization³, and the products were analysed by gel electrophoresis (Fig. 1c). An unamplified sample containing secret-message DNA yielded only a faint continuous smear (Fig. 1c, lane 2). In contrast, amplification of DNA microdots containing either 100, 10 or 1 copies of the secret-message DNA per haploid genome (lanes 3–5) each yielded a single product of the expected size (arrow). No such product was

detected using microdots containing either 0.1 (lane 6) or 0 (lane 7) copies per haploid genome, indicating a detection limit of about one secret-message DNA strand per haploid human genome. The amplified band in lane 4 of Fig. 1c (arrow) was excised, subcloned and sequenced. Use of the encryption key (Fig. 1b) to decode the resultant DNA sequence (Fig. 1d) yielded the encoded text, containing probably the most significant secret of the original microdot era: "June 6 invasion: Normandy" (Fig. 1d).

In preliminary experiments, microdots containing 100 copies of secret-message DNA per human haploid genome which had been attached using common adhesives to full stops in a printed letter, and posted through the US mail, yielded the correct PCR amplification product (data not shown). Our technique could therefore be used in a similar way to the original microdots: to enclose a secret message in an innocuous letter.

It should be possible to scale up the encoded message from the size of our simple example, perhaps by encoding a longer message in several smaller DNA strands. It should also be possible to use smaller microdots, which could be used for a variety of purposes, including cryptography and specific tagging of items of interest.

Catherine Taylor Clelland, Viviana Risca*, Carter Bancroft

Department of Physiology and Biophysics,
Mount Sinai School of Medicine, New York,
New York 10029, USA

e-mail: cbancro@smtplink.mssm.edu

*Present address: Paul D. Schreiber High School,
Port Washington, New York 11050, USA

1. Kahn, D. *The Codebreakers* (Scribner, New York, 1996).
2. Hoover, J. E. *Reader's Digest* **48**, 1–6 (April 1946).
3. Clayton, P. T. et al. *Arch. Dis. Child* **79**, 109–115 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Neurochemicals aid bee nestmate recognition

The theory of kin selection¹, which revolutionized the study of social behaviour, requires the discrimination of relatives from non-relatives. Many animals possess this ability, but the underlying neurobiological mechanisms have not been studied. Here we provide evidence for the neurochemical modulation of nestmate recognition: treatment with octopamine agonists improves the discrimination of related nestmates from unrelated non-nestmates in honeybees.

We used a modification of a laboratory assay² that measures the probability that a group of five-day-old, laboratory-reared adult worker honeybees (*Apis mellifera*) will

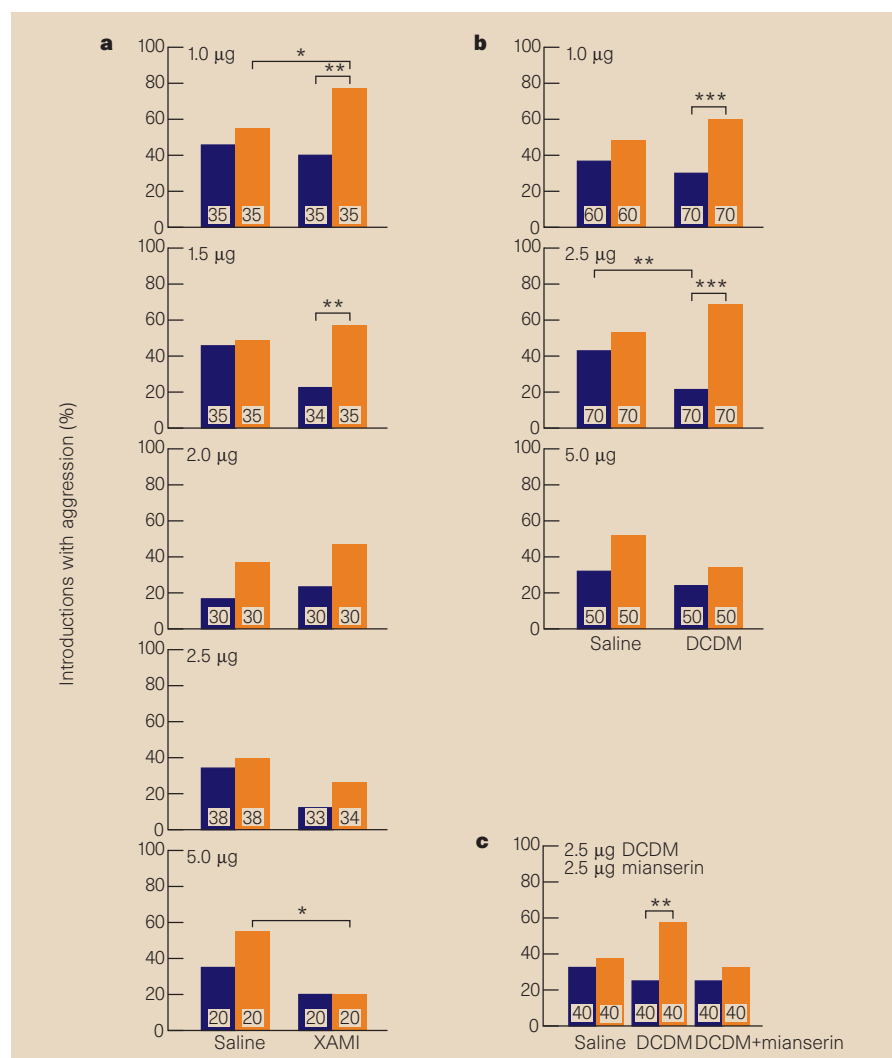


Figure 1 Effect of octopamine agonists on nestmate recognition in honeybees: **a**, XAMI; **b**, DCDM; **c**, DCDM plus mianserin (an octopamine antagonist). *P*-values are derived from two-way *G*-tests: **P* < 0.05; ***P* < 0.01; ****P* < 0.001. The number of introductions in each experiment is given in each bar. Purple bars, nestmate; orange bars, non-nestmate.

show aggression towards an introduced bee that is either a nestmate (from the group's natal colony) or a non-nestmate (from an unrelated colony). Group size was reduced from ten to three bees to make it possible for a single experimenter to inject all group members within a short period. Because the effects of octopamine injection are short-lived (about 60 min)³, we treated bees with more persistent octopamine agonists: either 2,3-xylylaminomethyl-2'-imidazoline (XAMI)⁴ or *N'*-(4-chloro-*o*-tolyl)-*N*-methylformamide (DCDM)⁵.

Bees given abdominal injections of 1.0 or 1.5 µg XAMI were significantly more likely to react aggressively towards non-nestmates than towards nestmates, but bees injected with saline were not (Fig. 1a). Comparisons with saline-injected bees suggest that the effect of XAMI was the aggregate result of two trends: decreased aggression towards nestmates and increased aggression towards non-nestmates. Only in one case was a trend significant by itself

(aggression towards non-nestmates, with a dose of 1.0 µg; this was significant only at the $\alpha = 0.05$ level, and many statistical tests were performed). Higher doses of XAMI (2.5 and 5.0 µg) did not influence nestmate recognition but did cause a significant transient impairment of locomotor activity (data not shown), which might have interfered with the expression of aggressive behaviour. These results are consistent with depressed responses to both nestmates and non-nestmates in bees treated with the two higher doses (Fig. 1a). Octopaminergic agonists have been shown to have an effect on locomotion in other species⁶.

To determine whether the results with XAMI reflected an octopaminergic process, we tested DCDM, an octopamine agonist from a different chemical family to XAMI. Results with DCDM were very similar to those with XAMI (Fig. 1b). Another indication of the specificity of the recognition effect is that it was eliminated when bees were treated with both DCDM and

mianserin (Fig. 1c), a potent antagonist at insect octopamine receptors *in vitro*^{4,5} and active in honeybees *in vivo*⁶.

The effective doses in this study were at least three orders of magnitude larger than endogenous levels in the brain⁸. Direct application to the brain might have allowed the use of lower doses, but would probably not have been compatible with a bioassay with unrestrained bees. Another possible side effect of this bioassay was that control bees did not discriminate between nestmates and non-nestmates as well as they normally do², perhaps because of the smaller group size.

Nestmate recognition is thought to require an individual to form a mental template based on learned (presumably olfactory) recognition cues of itself or its close relatives; the similarity of a conspecific's recognition cues to those represented by the template is then used to assess relatedness. We suggest that a higher octopamine concentration improves nestmate recognition by acting as a neuromodulator in the brain and increasing attention to relevant olfactory stimuli, so that fewer recognition mistakes are made. This is consistent with the finding that differences between treated and control bees resulted in both decreased aggression towards nestmates and increased aggression towards non-nestmates. It is also consistent with previous findings for bees that octopamine increases the rate of antennal scanning⁹, improves performance on an olfactory learning test⁴, and decreases the threshold of response for behavioural stimuli in other systems¹⁰. Octopamine levels in the bee brain increase with age⁸, and older bees are more defensive than younger bees. These results demonstrate that kin recognition, with widespread occurrence and compelling adaptive significance, offers promising experimental material for neuroethological analysis.

Gene E. Robinson*†, **Laura M. Heuser***, **Yves LeConte‡**, **Frederic Lenquette‡§**, **Robert M. Hollingworth¶**

*Department of Entomology, University of Illinois, Urbana, Illinois 61801, USA
e-mail: generobi@uiuc.edu

†Neuroscience Program, University of Illinois, Urbana, Illinois 61801, USA

‡INRA, Unité de Zoologie, Laboratoire de Biologie de L'Abeille, F-84914 Avignon Cedex 9, France

§CNRS, Laboratoire de Neurobiologie-Communication Chimique, F-13402 Marseille Cedex 9, France

¶Pesticide Research Center and Department of Entomology, Michigan State University, East Lansing, Michigan 48824, USA

1. Hamilton, W. D. J. *Theor. Biol.* **7**, 1–32 (1964).
2. Breed, M. D. *Anim. Behav.* **31**, 86–91 (1983).
3. Mercer, A. R. & Erber, J. J. *Comp. Physiol. A* **151**, 469–476 (1983).
4. Orr, N., Orr, G. L. & Hollingworth, R. M. *Insect Biochem. Mol. Biol.* **21**, 335–340 (1991).

5. Orr, N., Orr, G. L. & Hollingworth, R. M. *Insect Biochem. Mol. Biol.* **22**, 591–597 (1992).
6. Lund, A. E., Hollingworth, R. M. & Shankland, D. L. *Pestic. Biochem. Physiol.* **11**, 117–128 (1979).
7. Burrill, B. & Smith, B. H. *J. Insect Physiol.* **41**, 671–680 (1995).
8. Wagener-Hulme, C., Kuehn, J. C., Schulz, D. J. & Robinson, G. E. *J. Comp. Physiol. A* (in the press).
9. Pribbenow, B. & Erber, J. *Neurobiol. Learn. Mem.* **66**, 109–120 (1996).
10. Linn, C. E. & Roelofs, W. L. *Arch. Insect Biochem. Physiol.* **3**, 161–171 (1986).

Parameters for global ecosystem models

Tian *et al.*¹ have used their process-based ecosystem model to estimate the net CO₂ exchanges, also called net ecosystem productivity, for the years 1980–94. They deduced a large interannual variability ranging between –0.2 (from land to atmosphere) and +0.7 petagrams of carbon (Pg C) per year, the variability being mostly a function of soil moisture, which in turn is largely regulated by precipitation and temperature. These values were derived by including the modelled effects of increasing atmospheric levels of CO₂. The above numbers are the differences between net primary productivity and heterotrophic respiration. Over the given time period for the CO₂ feedback case, these values were 5.0 (±0.3) and 4.8 (±0.1) Pg C per year, respectively. The calculated net ecosystem productivity was thus a small fraction, between –4% and +14%, of the net primary productivity, with an average over the 15-year period of +4%.

We wish to point out that another carbon-release process, involving the emissions of isoprene (C₅H₈) and other volatile organic compounds (VOCs) from the Amazon vegetation to the atmosphere, must be taken into account. Guenther *et al.*² have estimated a global VOC emission of 1.15 Pg C per year, corresponding to an average global fraction of 2.4% of the net primary productivity. With the relatively larger VOC contributions from deciduous forests, this fraction will be of the order of 3% for tropical forests. The oxidation of isoprene to CO₂ does not occur promptly, but involves many reactions in which compounds such as CO and oxidized hydrocarbons are formed; these compounds have lifetimes of the order of days to a month (in the case of CO). The neglect of VOC emissions can lead to a substantial underestimate of the ultimate release of carbon from the tropical ecosystems to the atmosphere.

In our research in the forested regions of Surinam during March 1998 (refs 3,4), we used a proton-transfer-reaction mass spectrometer on an aircraft to detect VOCs. We measured not only high concentrations of short-lived (hours) isoprene averaging 2.8 nmol mol^{–1}, but also high concentrations of

up to 13.5 nmol mol^{–1} of various VOC compounds with a positive ion mass of 43 atomic mass units such as propan-2-ol, as well as formic acid and acetic acid with average molar ratios approaching 10 nmol mol^{–1}.

In agreement with Guenther *et al.*², we conclude that a proper forest-ecosystem carbon budget analysis must take into account the emissions of VOCs as well as of CO₂. Not only are direct emissions of VOCs from intact vegetation important, but leaf wounding is a significant additional source of VOCs⁵; herbivores consume (and thus wound) about 40% of tropical forest above-ground net foliar production⁶. Research is greatly needed to make better estimates of these emissions and the processes affecting them in order to expand the model of Tian *et al.*¹. This is also important for an understanding of the air chemistry above the tropical forests and its influence on global chemistry^{7,8}.

The occurrence of fires is an additional factor that must be taken into account. 'Dry' years, during which emissions of CO₂ are highest owing to low soil humidity¹, are also years of more extensive fires in the tropics, further increasing the interannual variability of carbon release.

Paul J. Crutzen*, **Ray Fall†**, **Ian Galbally‡**, **Werner Lindinger§**

*Max-Planck-Institute for Chemistry, PO Box 3060, 55020 Mainz, Germany
e-mail: air@mpch-mainz.mpg.de

†Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA

‡Division of Atmospheric Research, CSIRO, Private Bag 1, Aspendale, Victoria 3195, Australia
§Institute for Ion Physics, University of Innsbruck, A-6020 Innsbruck, Austria

1. Tian, H. *et al. Nature* **396**, 664–667 (1998).
2. Guenther, A. *et al. J. Geophys. Res.* **100**, 8873–8892 (1995).
3. Warneke, C. *et al. J. Atmos. Chem.* (submitted).
4. Williams, J. *et al. J. Atmos. Chem.* (submitted).
5. Kirstine, W. *et al. J. Geophys. Res.* **103**, 10605–10619 (1998).
6. Randerson, J. T. *et al. Global Biogeochem. Cycles* **10**, 585–602 (1996).
7. Houweling, S., Dentener, F. & Lelieveld, J. *J. Geophys. Res.* **103**, 10673–10696 (1998).
8. Wang, Y., Jacob, D. J. & Logan, J. A. *J. Geophys. Res.* **103**, 10757–10767 (1998).

We agree that the Terrestrial Ecosystem Model analyses by Tian *et al.*¹ of carbon fluxes of Amazonian ecosystems represent a methodological improvement compared with extrapolation from site-specific estimates², especially with regard to spatial resolution. However, the resolution of a model cannot be finer than that of the input data, and Tian *et al.* disregard one important group of ecosystems: peatlands. This is understandable, as the literature grossly underestimates the extent of peatlands in Amazonia. Our estimate is 150,000 km², ten times more than previously reported.

A global peat resource assessment³ erroneously implied a mire area of 15,000 km²

for the Amazon Basin, limited to Brazil. But western Amazonia harbours extensive swamps dominated by the palm *Mauritia flexuosa* L.f., as well as innumerable smaller mires along rivers, around lakes, in stream valleys, and in minor depressions within the rainforest. *Mauritia* swamps are estimated to cover 47,140 km² in Peru⁴, and vegetation maps suggest 8,000 km² for Colombia⁵ and 4,000 km² for Ecuador⁶. Our work indicates that peat strata in these swamps are often more than a metre thick. Considerable areas of such peatlands are also thought to exist in Brazil⁷.

Small mires within the rainforest are difficult to map, because they are indistinguishable in satellite images. However, on the basis of experience from 220 km of floristic transects in Peruvian, Colombian and Ecuadorian non-inundated rainforests, we estimate that peat deposits cover about 1% of their area, totalling almost 9,000 km². If the same proportion applies in Brazil, then it has more than 40,000 km² of undocumented peatlands intermingled with 'true' rainforest. In addition, Bolivia, Venezuela and the Guyanas are likely to harbour unreported mires. Hence, we estimate that there are around 150,000 km² of peatlands throughout Amazonia.

The average net ecosystem production of peatland ecosystems has remained positive for millennia, manifested by accumulation of peat. In nutrient-poor boreal mires, summertime (5 months) net ecosystem production can be 119 g C m⁻² (ref. 8), about three times the average (1980–94) annual value (about 42 g C m⁻², "climate with CO₂") reported by Tian *et al.*¹ for the Amazon Basin. However, peat deposits can also release considerable amounts of carbon. When the water-table is exceptionally low, summertime carbon emissions from boreal peatlands can be 83 g m⁻² (ref. 8), more than twice the highest annual value (about 40 g m⁻²) of Amazonia¹. Hardly any data exist on carbon fluxes of tropical peatlands. During drought, constantly high temperatures presumably render them strong carbon emitters. As soil moisture is apparently an important controller of carbon storage in Amazonia¹, incorporating such a response into the Terrestrial Ecosystem Model would further increase carbon emissions in El Niño years. The net ecosystem production of Amazonian peatlands is hard to estimate without ecological knowledge of the systems, but it might significantly affect the total carbon budget of the basin.

Finally, the soil organic carbon density (C_s) estimate⁹ used to validate the Terrestrial Ecosystem Model¹ was based on the RADAMBRASIL survey¹⁰, which excluded peat soils and covered only Brazil. When peatlands and other Amazonian countries are also included, C_s becomes close to 12 kg C m⁻², which is 30% greater than

the value obtained with the Terrestrial Ecosystem Model¹.

Leif Schulman, Kalle Ruokolainen,

Hanna Tuomisto

Department of Biology, Section of Ecology,
University of Turku, FIN-20014 Turku, Finland
e-mail: leif.schulman@utu.fi

1. Tian, H. *et al.* *Nature* **396**, 664–667 (1998).
2. Grace, J. *et al.* *Science* **270**, 778–780 (1995).
3. Lappalainen, E. (ed.) *Global Peat Resources* (Int. Peat Soc., Jyväskylä, Finland, 1996).
4. *ONERN Perfil Ambiental del Peru* (Oficina Nac. de Eval. de Rec. Naturales, Lima, 1986).
5. Etter, A. *Mapa general de Ecosistemas de Colombia* (Inst. de Inv. de Rec. Biol., Alexander von Humboldt, Bogotá, 1998).
6. Winkell, A., Zebrowski, C., Sourdat, M. & Zavgorodnyaya de Costales, S. in *Geografía Básica del Ecuador*, Tomo 4, Vol. 2 (Centro Ecuatoriano de Investigación Geográfica, Quito, 1997).
7. Suszczynski, E. F. in *Proc. 7th Int. Peat Congr., Dublin Vol. 1*, 468–492 (Int. Peat Soc., Jyväskylä, Finland, 1984).
8. Alm, J. *et al.* *Ecology* **80**, 161–174 (1999).
9. Moraes, J. L. *et al.* *Soil Sci. Soc. Am.* **59**, 244–247 (1995).
10. *Projeto RADAMBRASIL Levantamento De Recursos Naturais* Vols 3–20, 22, 25 & 26 (Dept Nac. Prod. Min., Rio de Janeiro, 1973–84).

Tian *et al.* reply — Our model-based analysis of the effects of interannual climate variability and increasing atmospheric CO₂ concentration on carbon storage in Amazonian ecosystems focused on CO₂ exchanges between the atmosphere and undisturbed forests and other upland ecosystems of the region¹. Crutzen *et al.* urge us to add the emissions of isoprene and other volatile organic compounds (VOCs) to our analysis. They argue that ignoring these emissions could lead to an overestimation of annual net carbon storage (net ecosystem production) in the Amazon Basin.

We did not include VOCs in our Terrestrial Ecosystem Model because not enough is known about their production, such as controls on rates, and tree species involved². The parameterization of CO₂ uptake (gross primary production) in the model is based on an estimate of the sum of net primary production and plant respiration, and does not include allocation of carbon to support the production of VOCs. Because we make no allowances in the model for CO₂ uptake by plants to support the production of VOCs, we make no allowances for emissions associated with VOCs. The estimates of net ecosystem production in our current version of the model are therefore independent of VOC emissions, and should not be corrected downwards for them.

However, the future development of the Terrestrial Ecosystem Model will certainly include the addition of VOCs because of their importance in tropospheric chemistry. Because the model is subject to mass balance constraints, we expect that our estimates of both gross primary production and net primary production will increase to accommodate the addition of VOC fluxes.

Schulman *et al.* suggest that we consider carbon fluxes between the atmosphere and peatlands in our calculations of net ecosystem production for the Amazon Basin.

They state that these ecosystems cover a large area, and they assume that tropical peatlands are likely to be at least as responsive to climate changes as their boreal counterparts. From our review of the literature on the areal extent of peatlands in Amazonia^{3,4}, we conclude that Schulman *et al.*'s estimate of peatland area in the basin, 150,000 km², is reasonable. If we combine the total peatland area of Amazonia with the boreal peatland flux rates cited by Schulman *et al.*, the resulting basin-wide fluxes are small. A carbon storage rate (positive net ecosystem production) of 119 g C m⁻² yr⁻¹ translates to an annual basin-wide storage of about 0.02 Pg C, and a carbon loss rate (negative net ecosystem production) of 83 g C m⁻² yr⁻¹ translates to an annual basin-wide release of about 0.01 Pg C. For interannual climate variability to have a significant effect on the net ecosystem production of Amazonia through peatlands, these ecosystems would have to be much more sensitive than boreal peatlands to climate shifts.

There is evidence that this is not the case. The literature on peatlands in warm climates indicates that, because of the poor quality of their organic matter, decomposition rates in these ecosystems are low under both aerobic and anaerobic conditions⁵. This is not true for boreal peatlands, where the low temperatures that prevail for most of the year slow the decay of plant litter. Slow decay leads to a build-up of relatively high-quality organic matter that decomposes rapidly under warmer and drier conditions. Because tropical peatlands may be less sensitive than boreal peatlands to interannual climate variability, and because the area of peatlands is relatively small in Amazonia, we conclude that the net ecosystem production of the Amazon Basin is little influenced by the effects of year-to-year variability on carbon storage in its peatlands.

Hanqin Tian*, Jerry M. Melillo*, David W. Kicklighter*, A. David McGuire†, Berrien Moore III‡, Charles J. Vörösmarty‡

*The Ecosystems Center,
Marine Biological Laboratory,
Woods Hole, Massachusetts 02543, USA

†US Geological Survey,
Cooperative Fish and Wildlife Research Unit,
University of Alaska, Fairbanks, Alaska 99775, USA

‡Institute for the Study of Earth, Oceans and Space,
University of New Hampshire,
New Hampshire 03824, USA

1. Tian, H. *et al.* *Nature* **396**, 664–667 (1998).
2. Lerdau, M., Guenther, A. & Monson, R. *Bioscience* **47**, 373–383 (1997).
3. Junk, W. J. in *Mires: Swamp, Bog, Fen and Moor* (ed. Gore, A. J. P.) 269–294 (Elsevier, Amsterdam, 1983).
4. Junk, W. J. in *Transport of Carbon and Minerals in Major World Rivers* (eds Degens, E. T. H., Kempe, S. & Herrera, R.) 267–283 (Mitt. Geol.-Palaont. Inst. Univ. Hamburg, SCOPE/UNEP Sonderbd, Hamburg, 1985).
5. Bridgman, S. D. & Richardson, C. J. *Soil Biol. Biochem.* **24**, 1089–1099 (1992).

Apes and their place in our world

Brutal Kinship

by Michael Nichols and Jane Goodall
Aperture: 1999, 127 pp. \$25, £15.95

Mary Midgley

At first glance, *Brutal Kinship* looks like a simple picture-book: a collection of magnificent, thoughtful photographs of chimpanzees in various situations. It is that, but it is also something more. The text is mostly by Michael Nichols, who took the photographs, but partly also by Jane Goodall, who accompanied him. They describe the life of the apes in the wild and contrast it, quietly and factually, with the details of the apes' conditions in various kinds of captivity — zoos, circuses, laboratories, breeding facilities, holding stations, backyards in African villages and private homes. Once grasped, this contrast cannot fail to be shocking. The writers do not ham up the facts, because they do not need to.

"As soon as I became aware of the plight of the chimpanzees at the hand of *Homo sapiens*, I felt that I had to make a strong statement and effect change," says Nichols. He and Goodall note that the apes' similarity to us in intelligence and genetic structure allows us to learn from them about human thought, speech, sociality and disease. "Yet we refuse to extend to them even the most basic rights that beings with their intellectual and emotional qualities deserve."

The authors hope to arouse awareness of what they call "our moral myopia". They found, for instance, that staff at medical laboratories genuinely cared about the chimpanzees in their keeping while having little control over their conditions. "But this does not excuse them or us of our responsibility," say the authors. "If we can see that our treatment of chimpanzees has been and is wrong, then we will have truly evolved."

In short, something is happening which has occurred repeatedly in human history. The institutions involved are anachronistic; they were formed when people had a different set of beliefs about the relation between humans and other animals from the one we have today. Institutions, however, persist by their own inertia. Their persistence leads to grating moral conflict.

The central change involved is, of course, one within science itself. Today, our understanding of our evolutionary continuity with the rest of the natural world calls on us to attend to similarities, to understand humans as a kind of animal and other animals as akin to humans. From that perspective, the burden of proof tends to be on anyone who asserts that animals are radically alien to us.

But during the formative period of modern science, things were very different. The reigning Christian perspective of the time



Rival claims: a chimpanzee responds to Jane Goodall's pant-grunts of greeting by refusing to return to his beach photographer owner.

showed humans as radically distinct from the rest of nature, since they were made in God's image and were the only creatures possessing souls. It is unlucky that the world-view of scientists at that time was so deeply influenced by Descartes, who hardened and expanded this mind-body dualism, equating soul with consciousness, and ruling that non-human animals, being soulless, were literally unfeeling automata.

Descartes' dualism was widely welcomed because it helped to define and establish a clear realm for physics. Dualism therefore pervaded the idea of a 'scientific' outlook which was so important to Enlightenment thought. It carried with it the notion of intelligent human observers as essentially magisterial beings, minds standing outside and above the brute physical world that they studied. From this perspective, attempts to relate human experience to that of other animals were considered illicit, a sentimental error stigmatized as anthropomorphism. Scientific thinkers, even when they did not think animals were actually unconscious, commonly took these lives to be so distant and inferior that they could not possibly concern humans.

Two things, however, went wrong with this dualism. First, the idea of a separate soul went out of fashion, taking with it the disembodied observer. Second, as knowledge of biology grew, the wide gap that was supposed to divide us from all other creatures began gradually to fill up with intermediate terms. The discovery of the great apes themselves was, of course, a crucial stage in this process, and the Darwinian revolution finally gave it shape. That tremendous change, if it is taken seriously, surely means that what we think of as a 'scientific attitude' today is much more

likely to be one that sees humans as an integral part of the natural world than the seventeenth-century one which placed them outside it.

Unavoidably, this change has practical consequences. It questions the rigid barrier that was held to exclude other creatures totally from our consideration. Goodall asks: "How should we relate to beings who look into mirrors and see themselves as individuals, who mourn companions and may die of grief, who have a consciousness of 'self'? Don't they deserve to be treated with the same sort of consideration we accord to other highly sensitive beings — ourselves?" Her question comes from inside science, not outside it. It does not arise from ignorance, but from her thorough, well-documented understanding of the apes, gathered through scientific work with them over 40 years.

She speaks, in fact, for many scientists who have moved on from Descartes' world to Darwin's, scientists who no longer think that there is anything unscientific about being humane. For them, the paradox that Nichols mentions — that the apes' likeness to us both strengthens their claims on us and, at present, subjects them to extra ordeals — is profoundly disturbing. The question of how we can act on these new insights, how to make our institutions less unsuitable to our moral perceptions, is a hard one.

This book carefully describes the practical problems and explains many ingenious devices by which dedicated people are working to resolve them. It calls for our help in these efforts. And that call is surely entirely in accord with the spirit of science today. □

Mary Midgley is at 1a Collingwood Terrace, Newcastle-upon-Tyne NE2 2JP, UK.

Telling it like it was

Deep Time: How Humanity Communicates Across Millennia

by Gregory Benford
Avon: 1999. 225 pp. \$20

Peter T. Landsberg

To transfer information at some point in time to a period that could be later by millennia is not an easy proposition, but such transfers have a long history. Think of the pyramids and the various structures that are known as the Seven Wonders of the World. Apart from the Cheops pyramid in Giza, Egypt, most did not survive as long as their creators had hoped. But the point is that the ancient structures that did survive serve a useful purpose: they furnish us with a guide to the sort of life lived many centuries ago. And, of course, they gave reigning emperors some assurance that they would be comfortable in their afterlife. The cave paintings in the Lascaux cave, on the other hand, may represent art for art's sake, or may also have had a magical purpose. We just do not know.

In the present era, it is not the wish for comfort in the hereafter, or the desire to exercise a religious impulse, that causes us to signal to the future. Our motivation is scientific fervour. The 1977 Voyager missions to the outer planets, for example, contained a phonograph record of the shouts and songs of humanity (plus instructions on how to use it). This reminds one of the 1971 Pioneer mission, which carried a plaque showing two nude humans greeting the infinite with a wave. Far more complicated ideas had also been considered by the scientists of the various space agencies. For example, the finders of a payload would have been able to date a launch to within a million years or so had it provided a current view of the Andromeda galaxy. The reason? The galaxies rotate, though rather slowly, and the observed rotation would have revealed roughly how much time had passed.

Gregory Benford notes that the 1997 Pathfinder spacecraft deposited the names of members of the Planetary Society on Mars — probably not essential information for future cosmic travellers. Indeed, as Benford says, “as a projection of pure vanity, it resembles the International Star Registry, which sells people certificates stating what stars have been named after them”.

“Deep time”, a period of the order of a millennium or so introduced by the author, is also relevant for those who wish to save our biological diversity. Benford relates how he wrote a paper in 1992 in which he proposed to take representatives of all threatened species and suspend them in liquid nitrogen for the indefinite future. This paper was



From there to here: our desire to leave a marker for the future existed back in ancient Egypt.

rejected by *Science* and *Nature*, but eventually caused considerable discussion after its publication in the *Proceedings of the National Academy of Sciences*. The idea is to go beyond “the piecemeal strategies of seed banks, of germ plasma”, to make up an, admittedly incomplete, collection of threatened species. There is a full discussion of this suggestion in a chapter called “The library of life”. The problem of showing precisely where radioactive waste is buried is discussed in a chapter entitled “Ten thousand years of solitude”. The markers should presumably survive thousands of years, but there is a case also, as pointed out by the author, for not marking them at all, so as to prevent future theft.

Deep Time offers an unusual combination of ideas that give food for thought, even though some seem rather far-fetched. In a future edition, Benford might improve the organization of the text, for example by giving a table of dates and other details of the satellite launchings that are discussed. Also, a list of markers — ‘headstones’ — actually used to indicate burial sites of radioactive waste would help. Further, the author’s colleague Frank Drake, who is mentioned repeatedly, is presumably connected with the well-known equation that gives a rough estimate of the number of advanced civilizations in a galaxy. This might be pointed out.

Benford is a professor of physics at the University of California at Irvine and helped to design the message that flew on the 1997 Cassini mission to Saturn. He has also worked on the long-term marking of US nuclear-waste sites. So the book is worth reading for its insider’s account of the design of markers that tell the story of humanity, and of the Earth, for the benefit of future cosmic travellers. It is also useful for the details it

gives of the plans to keep alive our biodiversity, and for its discussion of the storage of nuclear waste. □

Peter T. Landsberg is in the Department of Mathematics, University of Southampton, Southampton SO9 5NH, UK. His book *Seeking Ultimates: An Intuitive Guide to Physics* will be published by the Institute of Physics this autumn.

Explosion of interest

Melting the Earth: The History of Ideas on Volcanic Eruptions

by Haraldur Sigurdsson
Oxford University Press: 1999. 260 pp.
\$30, £18.50 (pbk)

Martin Rudwick

Volcanoes can hardly be ignored, least of all by those who live near them, and their often violent eruptions have duly been recorded throughout literate human history. One of the earliest to be described in some detail was the eruption of Vesuvius in AD 79, which destroyed the towns of Pompeii and Herculaneum. Yet, by preserving the buildings and their contents, that eruption has also given historians and archaeologists an almost unique insight into the everyday life of a Roman province. Twenty years ago, the volcanologist Haraldur Sigurdsson became intrigued by what the volcanic debris that had buried the towns could reveal about the eruption itself. That scientific research, aided by Pliny’s famous eye-witness account, led to this historical survey of ideas about volcanic eruptions from antiquity to the present day.

Although Sigurdsson writes as a scientist

rather than as an historian, he is familiar with some of the historical literature and, more importantly, has read for himself many of the historical sources that he quotes or summarizes. His account is illustrated with many contemporary pictures of eruptions, some of which will be unfamiliar even to historians of the Earth sciences. He also has the inestimable advantage of having grown up in the volcanic landscapes of Iceland.

Sigurdsson's account combines two distinct kinds of history. The first, which predominates in the earlier chapters, is a history of the eruptions themselves and their human impact. He summarizes the long-running debate about the relation between the catastrophic eruption of Thera in the Aegean in the seventeenth century BC, and the nearly contemporary collapse of the Bronze Age Minoan civilization. He also gives a fascinating reconstruction of the complex eruption of Vesuvius in AD 79.

The second kind of history, which almost takes over in the later part of the book, is an account of the ideas of savants (who in more recent times can properly be called scien-

tists) about the causes of volcanic activity. Here Sigurdsson wants to shift historical attention away from those, such as Nicolas Desmarest in eighteenth-century France, who first recognized extinct volcanoes in regions far from active ones, and thereby showed that volcanism was no mere minor or local phenomenon. Instead, he highlights those he regards as the "true heroes" of his story, nineteenth-century geophysicists, such as William Hopkins.

This is fair enough, but Sigurdsson fails to recognize that his own interest in the causes of volcanic activity was, for savants of earlier generations, only one of many features of volcanoes that deserved close scientific study. For them, the descriptive and classificatory 'natural history' of eruptions was just as important as any 'natural philosophy' or causal theorizing about them. And the former often had greater prestige precisely because it could be more rigorous and not as speculative as the latter.

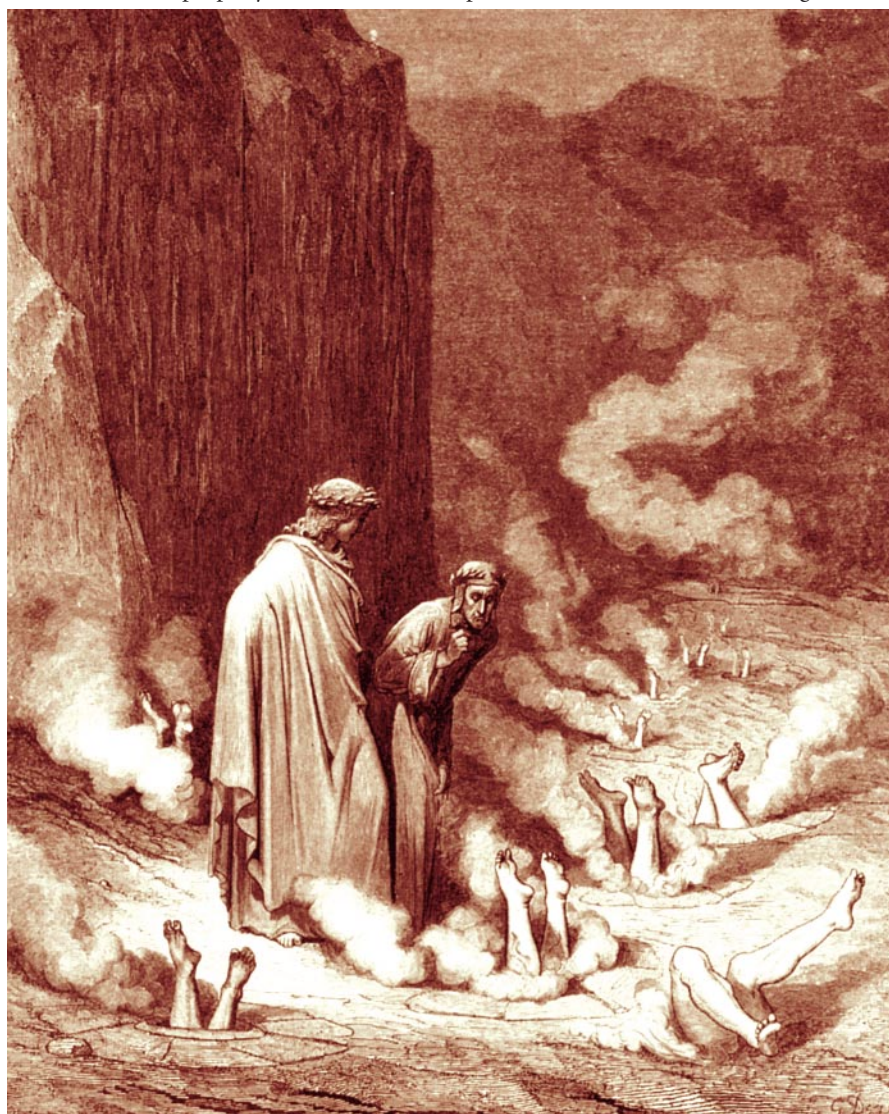
Furthermore, even within the tradition of causal theorizing, there were many other problems besides the one that Sigurdsson,

from his modern perspective, regards as the most fundamental and uses for his title. The mechanism by which rocks are melted within the Earth, to form the magma that reaches the surface as volcanic lava, was indeed a long-standing puzzle. But so, for example, was the very fluidity of lavas and the origin of their generally crystalline structure, and the reasons for the striking variations in the behaviour of different volcanoes.

Sigurdsson does in fact deal well with the history of some of these problems, and he also gives a reliable account of the fortunes of various ideas about the ultimate source of the heat that locally melts the Earth into magma. He takes the story through the discovery of radioactivity a century ago, which provided a previously unsuspected source of heat, and through the development of plate-tectonic theory in the 1960s. And he ends in the age of space exploration with a brief allusion to the extraterrestrial volcanoes discovered recently on Jupiter's satellite Io.

All in all, Sigurdsson has provided an attractive and readable account of the history of ideas about volcanoes, and it should do much to curb what he describes as scientists' frequent dismissal of their own past, "an arrogance that is built on ignorance". □

Martin Rudwick is in the Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, UK.



Hell on Earth: Virgil in Dante's *Inferno*, inspired by bubbling mud pits in a volcanic region near Naples.

From tulips to electric cars

Arbres de Pierre: La Croissance Fractale de la Matière

by Vincent Fleury

Flammarion: 1999. 334 pp. FF170

Patrick Tabeling

The beauty and complexity of the forms found in nature have always intrigued and challenged physicists. Where do they come from? Are there general, underlying, perhaps simple, mechanisms that lead to the formation of a snowflake or a tulip? These questions form the theme of Vincent Fleury's *Arbres de Pierre*, which describes morphogenesis from historical, physical and aesthetic perspectives.

The first chapters give a well-documented historical background to the subject. We read how René-Antoine Réaumur, inventor of the thermometric scale that bears his name and a pioneer in the days before morphogenesis was recognized as a subject, was concerned by the inability at the time to explain the origin of the shapes of stones. We learn how Johannes Kepler proposed a mechanism for the formation of snowflakes, taking into account their sixfold symmetry. We wander through previous centuries,

following a tortuous path along with the issues that were raised, eventually jumping 200 years to the latest developments which came with the marked progress in our understanding of nonlinear phenomena.

Throughout, Fleury avoids giving any equations. Would equations frighten non-physicist readers? This is a reasonable concern, but it tends, in my view, to hamper the presentation of several important aspects of the subject. For instance, crystal growth, which can be succinctly described by the diffusion equation, is here arduously represented in terms of a drunk's random way of walking. The outstanding progress made in the past decade on growth phenomena, which relies mostly on subtle mathematics, is laboriously presented. The reader may be reminded here of Richard Feynman's remark that the role of formulae is to shorten the description of laws that are otherwise cumbersome to state. This policy of rejecting any formal material in favour of lengthy discussions nonetheless provides the reader with a solid, bottom-up understanding of growth phenomena.

The breath of inspiration flows poetically through the book, but even so, the industrial applications of morphogenesis are not overlooked. The role of dendritic growth in electrochemical battery recycling, and why electrochemical growth on electrodes, whether fractal or dendritic, is a major obstacle to the development of electric cars, are clearly explained. The discussion of dendrite formation in the glass industry is also interesting. I would have expected more space to be given to dendritic and fractal crystal growth in metallurgy — perhaps the most important field for research, whether applied or theoretical. The author is clearly not attempting exhaustive coverage of his subject. However, the chapter on plants is probably the best a physicist can say on subjects such as phyllotaxy or pattern formation in biology.

Those most likely to enjoy this book are physicists — researchers, teachers or students — interested in the formation of morphologies. But others with a background in physics might also find it interesting. The book offers physical insights, enlightening historical perspectives and up-to-date presentation of discoveries and concepts in morphogenesis. It is sometimes said that research problems are simply those that strike the human mind. Morphogenesis is one of them, and this book is a vivid and delightful account of it. □

Patrick Tabeling is at the Laboratoire de Physique Statistique, ENS, 24 rue Lhomond, 75231 Paris, France.

More, morphologically speaking Universalités et Fractales: Jeux d'Enfant ou Délits d'Initié?

by Bernard Sapoval
Flammarion, FF160 (pbk)



Size is relative: the giant millipede *Spirostreptus giganteus*.

Stumbling about the world of little things

Waiting for Aphrodite

by Sue Hubbell
Houghton Mifflin: 1999. 242 pp. \$24

Martin Wells

For the past 40 years I have been trying to persuade students to do just what Sue Hubbell has done. I give them an essay title, tell them to go to the libraries and find out everything they can about it, ask advice, discuss it with their colleagues and then write an account for an interested layperson, in good English, with the minimum of scientific jargon. I point out, hoping to appeal to venality where an appeal to scholarship may fail, that there is money in this. You can flog the output as well as passing your exams. I tell them about David Attenborough, who did the same courses as I did at Cambridge two years ahead of me and then went on to base much of his glorious "Life on Earth" series on the biology we were both taught.

A pushover, I say, an example of the immediate cash value of education.

This is a lie. Whatever else it is, it isn't a pushover. You have to be rather good at it to make it stick. Hubbell is very good at it. In this random walk among the invertebrates, she picks her essay subjects cleverly, selecting animals, like millipedes and woodlice, sea urchins and spiders, that we all know but know very little about. She does her library work, tracks down the experts — she says that experts, particularly those in the less fashionable areas of biology, are always willing to talk about their specialities (I suspect she has charm as well as the capacity to do her homework) — and then writes up some very

interesting revelations. For example, I now know how a parasite of woodlice manages to include starlings in its life history, despite the normally cryptic habits of woodlice; about the regulation of sea-urchin fisheries, suffering now on the US Eastern seaboard because the Japanese love eating the gonads. I know about monoculture and the inadequate flight range of most species of bee; and why the generic name *Aphrodita*, far from being a tribute to the golden-haired goddess, is the result of a tasteless joke by Linnaeus based on the animal's common name, the 'sea mouse', and 'mouse', which just happens also to be a Northern European term for a woman's sexual parts. The man was a Swede. One lives and learns.

I hadn't read any of Hubbell's five other books before following this random walk. I shall now. She obviously gets fun out of biology, makes life more interesting

for other people as well as herself, and — as far as I can detect — doesn't make mistakes. There is a bibliography at the end of each chapter which very sensibly includes elementary textbooks as well as more specialized material, just in case you want to check on her. There are occasional folksy bits (well, Americans are like that, aren't they?) that I am not too keen on, although I guess they help fill in the background. Actually, I do like to know something about the authors I read; the stuff on the dust-jackets rarely tells one anything significant. And, unlike Hubbell, good countrywoman as she is, I've always been a dog-hater — they rush about and disturb the wildlife. But all is forgiven, because this is top-level popularization and I wish we had more of it. □

Martin Wells is in the Department of Zoology, University of Cambridge, Cambridge CB2 1TN, UK.



Marine viruses and their biogeochemical and ecological effects

Jed A. Fuhrman

Viruses are the most common biological agents in the sea, typically numbering ten billion per litre. They probably infect all organisms, can undergo rapid decay and replenishment, and influence many biogeochemical and ecological processes, including nutrient cycling, system respiration, particle size-distributions and sinking rates, bacterial and algal biodiversity and species distributions, algal bloom control, dimethyl sulphide formation and genetic transfer. Newly developed fluorescence and molecular techniques leave the field poised to make significant advances towards evaluating and quantifying such effects.

The importance of microbial processes in the sea has become increasingly apparent over the past two decades. Although heterotrophic microbes (see Box 1) in the water column used to be generally ignored, it is now clear that a large proportion (often half or more) of the total flux of matter and energy in marine food webs passes through such organisms by means of dissolved organic matter^{1,2}. Heterotrophic prokaryotes are abundant¹, representing a significant proportion of the biomass in the euphotic zone (the surface waters), and an even larger proportion in deeper waters³. The concept of the microbial loop in marine food webs has been developed and refined in recent years, but until a decade ago, viruses were ignored in such studies. However, a series of reports then showed that viruses are not only remarkably abundant in marine plankton, but are also likely to be significant agents in the control of bacteria and phytoplankton^{4–7}.

Viruses are small particles, usually about 20–200 nm long, consisting of genetic material (DNA or RNA, single- or double-stranded) surrounded by a protein coat (some also have lipids). They have no intrinsic metabolism and function only as parasites through the cellular machinery of a host organism. In theory, all cellular organisms are susceptible to infection, often by more than one type of virus, implying that viruses are probably the most diverse creatures on Earth. A given type of virus usually has a restricted range of hosts—often a single species, although some viruses infect only certain subspecies, whereas others may infect more than one related species or even genus⁸. Viruses contact their hosts by passive diffusion and use their host's exposed cellular structures, often transporter proteins, as attachment and entry points to the cell.

There are three basic types of virus reproduction. (1) Lytic infection: the virus attaches to a host cell and injects its nucleic acid into the cell, directing the host to produce numerous progeny viruses; these are released by fatal bursting of the cell, allowing the cycle to begin again. (2) Chronic infection: the progeny virus release is non-lethal and the host cell releases them by extrusion or budding over several generations. (3) Lysogeny: the nucleic acid of the viral genome becomes part of the genome of the host cell, and reproduces as genetic material (called a prophage or provirus) in the host cell line. In this case, an induction event, such as stress to the host, can trigger a switch to lytic infection.

A less well-defined type of virus–host interaction is termed pseudolysogeny, in which the viral nucleic acid may remain within a host cell for some time (possibly for a few generations⁹) before lysis, or cell destruction, occurs. Pseudolysogeny may be related to host starvation, in which the virus adopts an inactive state, unable to initiate viral gene expression owing to the low energy state

of the cell; normal viral activity returns when the cell is fed¹⁰. Alternatively, pseudolysogeny may be regarded as a transient state of host immunity, apparently induced by an immunizing agent (perhaps a polysaccharide depolymerase) released from infected cells, and helping to foster coexistence of host and virus^{11,12}. Viruses or virus-like particles may also be involved in killing cells by mechanisms that do not result in virus reproduction¹³.

Here I synthesize the accumulated evidence regarding the nature of marine viruses and their ecological and biogeochemical effects. The principal conclusion is that viruses can exert significant control on marine bacterial and phytoplankton communities, with respect to both biological production and species composition, influencing the pathways of matter and energy transfer in the system.

Virus abundance

Viruses infecting specific marine organisms have been studied¹⁴ for several decades, initially from studies of pure cultures that focused on organisms rather than ecological systems; but viruses were not regarded as quantitatively important components of marine food webs until they were shown by direct counts to be highly abundant. The small size of viruses renders them invisible to ordinary light microscopy, and although they can be visualized by transmission electron microscopy (TEM), special procedures are required to concentrate them from sea water.

The ultracentrifugation and TEM method that is commonly used to collect and visualize marine viruses and prokaryotes is adapted only slightly from the original technique published¹⁵ in 1949—had this technique been applied then to natural waters, the discovery of highly abundant prokaryotes and viruses in aquatic habitats would have been pushed forward by about 25 and 40 years, respectively. This could have redirected the development of modern biological oceanography and limnology, which have only recently moved significantly towards microbiological studies.

The first reports of high viral abundance, exceeding the typical bacterial abundance of 10^9 per litre (refs 4–6), awakened interest in this topic. Many subsequent studies^{16–25} have shown that viruses are consistently the most abundant biological entities in the sea—nearshore and offshore, tropical to polar, sea surface to sea floor, and in sea ice and sediment pore water. Viral abundances are typically 10^{10} per litre in surface waters (about 5–25 times the bacterial abundance), and follow the same general abundance patterns as bacteria. These patterns include a decrease of about one order of magnitude between rich coastal waters and oligotrophic (nutrient poor) open ocean, a decrease of between five- and tenfold from the euphotic zone to the upper midwaters (for example, 500 m depth), and a further decrease several-fold to

Box 1 Glossary of marine organisms

Archaea. One of the two groups of *prokaryotes*, whose cultivated members are often methane-producing or tolerant to unusually high temperature or salinity. Uncultivated archaea, such as those from the deep sea, may have other physiologies.

Autotrophs. Organisms that use CO₂ as their source of carbon. (All green plants are autotrophs.) See also *heterotrophs*.

Bacteria. One of the two groups of *prokaryotes*; also used as a generic term describing organisms that appear to be *prokaryotic* but are otherwise unidentified (some may be *Archaea*).

(*Bacteriophage*s). Viruses that infect bacteria.

Cyanobacteria. Type of *bacteria* that contain chlorophyll *a* and undergo photosynthesis, generating oxygen. This phylogenetic group includes the marine *prochlorophytes*.

Cyanophages. Viruses that infect *cyanobacteria*.

Eubacteria. Another name for the *Bacteria* group.

Eukaryotes. Organisms with membrane-bound nuclei; all animals and higher plants are eukaryotes. (See also *protists*.)

Eukaryotic algae. Photosynthetic *protists*.

Flagellate. A type of *protist* that moves by beating flagella (long, strand-like structures present on the surfaces of some cells).

Heterotrophs. Organisms that derive energy from preformed organic matter. (All animals are heterotrophs.) See also *autotrophs*.

Metazoa. Multicellular animals.

Microbe. Microscopic organism, occurring as a single cell or simple colony.

Phytoplankton. Photosynthetic plankton, including *cyanobacteria* and *eukaryotic algae*.

Prochlorophytes. Small photosynthesizing *bacteria* that possess divinyl chlorophyll *a* and *b* (see also *cyanobacteria*).

Prokaryotes. Cells without membrane-bound nuclei. They include two broadly different groups, the *Bacteria* and *Archaea*.

Protist. A type of *eukaryote* that is single-celled, or lives in simple colonies of single cells. Protists are generally larger and less abundant than *bacteria*. (See also *flagellates*, *eukaryotic algae* and *protozoa*).

Protozoa. Old name for non-photosynthetic *protists*. (Protozoa are distinct from the *metazoa*.)

Zooplankton. Eukaryotic (animal) plankton. Macrozooplankton are visible to the naked eye and feed on the largest prey organisms. Microzooplankton are just below the visible size range (~20–200 µm), and may include very small *metazoa*, animal larvae and large *protists*. Nanozooplankton include the smallest *protists* (~2–20 µm) and eat the smallest prey, such as *bacteria*.

abyssal depths. As occurs with bacteria, sea ice is highly enriched in viruses compared with the water beneath it²¹, and sediment pore waters are highly enriched compared with overlying water^{19,24}.

Viral abundances are dynamic, being particularly responsive to changes in ecological conditions such as algal blooms^{7,26}—this provides strong evidence that viruses are active members of the community rather than inert particles. Pronounced fluctuations²⁷ in virus abundance over timescales of minutes to hours may be indicative of synchronized host-cell lysis and rapid degradation of a large portion of the progeny viruses. On scales of hundreds of kilometres, virus abundance is usually strongly correlated to bacterial abundance and less so to particulate chlorophyll *a* (refs 18–20), indicating that most marine viruses infect bacteria. The term ‘bacteria’ refers here and throughout to prokaryotes in general because most studies do not distinguish true Bacteria (also called Eubacteria) from Archaea (see Box 1); in some environments (particularly the deep sea), half of the prokaryotes may belong to the Archaea²⁸.

It is possible to use epifluorescence microscopy with nucleic acid stains such as 4,6-diamidino-2-phenylindole (DAPI), YoPro and SYBR Green to count viruses accurately, rapidly and inexpensively

compared with TEM^{25,29–32}. Viruses stained with SYBR Green can be observed (Fig. 1) and counted by ordinary epifluorescence microscopy in the field (for example, on board ship) within an hour of sampling; the counts obtained²⁵ are similar or slightly higher (by ~1–1.5 times) than those obtained by TEM. Viruses stained with SYBR Green may also be counted by flow cytometry³³, further expediting aquatic virus studies; again the counts are higher than TEM values.

The reasons for the higher fluorescence counts compared with those obtained from TEM preparations are uncertain, but fluorescence counts may include viruses that are otherwise obscured by other dark-stained particles in TEM preparations; fluorescence also enables counts of filamentous viruses or other types that are not recognized easily by TEM. The new stains that permit fluorescence imaging or flow cytometric detection of a diverse array of viruses may prove useful in biomedical studies, representing a spin-off of marine virus research.

Beyond simply counting viruses, marine viral diversity has also been examined by morphology and size distribution; cultures and natural samples show a broad array of shapes and sizes³⁴. A new technique, pulsed field electrophoresis^{35,36}, has revealed the size diversity of viral genomes extracted directly from marine samples. The method provides a minimum estimate of the diversity of the most abundant viruses. There are typically 15–40 visibly distinct genome sizes in a given sample, and the composition is variable in space and dynamic over seasons.

Viral activities

One of the initial reasons that biological oceanographers looked for abundant viruses was that there seemed to be excess bacterial production that had a missing sink, and viruses seemed good candidates for that sink². Viruses can limit bacterial abundance to several orders of magnitude below the resource-limited level, demonstrating a potential for control⁹. As soon as it was learned that viruses have consistently high abundances in sea water, it was hypothesized that these viruses must be infecting native marine microbes at significant rates in order to persist⁵.

Microbes, particularly prokaryotes, seemed to be the major host candidates because they are by far the most abundant potential hosts (~10⁹ marine bacteria per litre)—viruses reach their hosts by random diffusion and interception, with contact proportional to host abundance and size, so that a virus is much more likely to contact a bacterium than a protist or metazoan³⁷. The relationship between viruses and bacteria seems corroborated by field data showing that bacterial abundance is the best predictor of virus abundance, explaining about 67% of the variance¹⁸. However, given the presumed host specificity of viruses and the largely unknown species diversity and composition of marine microbial assemblages²⁸, quantitative evaluation of contact between marine viruses and hosts still involves much guesswork.

In the decade since the discovery of high viral abundance, several lines of evidence have converged to the conclusion that viruses are significant agents in both the mortality of aquatic microbes and in the structuring of aquatic communities. That evidence is summarized below.

Infection frequency. Direct evidence that viruses infect native bacteria and cyanobacteria originally came from TEM studies of thin-sectioned plankton, in which the assembled viruses were observed in about 0.8–4.3% of the cells from a variety of marine habitats: high-nutrient coastal waters, oligotrophic open ocean⁶ and sinking material collected in offshore sediment traps³⁸. Similar percentages are typical in freshwater samples³⁹, although some freshwater work has found lower percentages, occasionally undetectable²⁶.

As viruses are visually detectable inside hosts only at the last stage of lytic infection, the visibly infected cells represent only a small proportion of the total number of infected cells. A model based

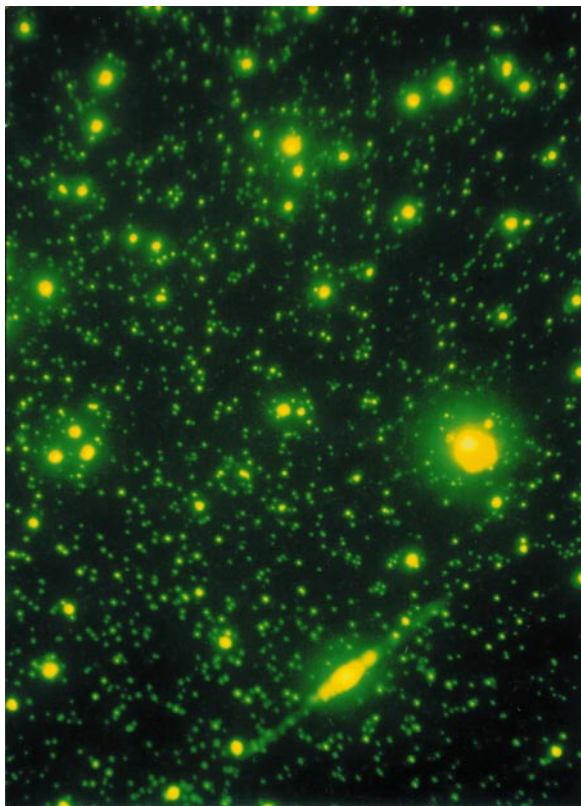


Figure 1 Fluorescence imaging of marine viruses. Epifluorescence photomicrograph of viruses (small, numerous green dots) and prokaryotic cells (rarer, larger green dots), collected from 5 m depth in the San Pedro Channel near Santa Catalina Island, California, on 29 May 1998 and prepared with SYBR Green I stain²⁶. Photographed on Kodak Ektachrome 400 film, 30 s exposure. The two largest stained objects are a pennate diatom (oblong) and an unidentified protist.

upon culture studies^{6,40} was used to calculate the total mortality from visibly infected cells. The model's predictions indicated that the apparent percentage infected should be multiplied by a conservative factor of about 10 (but ranging from 7.4 to 14.1) to estimate the total percentage of mortality due to infection. This analysis indicates that viruses are responsible for about 8–43% of the bacterial mortality in the sites mentioned above. However, this may be an overestimate, because an infected cell may be grazed before it lyses⁴¹.

Some studies^{39,42–44} simplify the examination of infected cells by looking through intact cells, rather than preparing thin sections. The former procedure is much faster, and requires only a few millilitres of sample. A slightly higher percentage of visibly infected cells is often found compared with the thin-section method⁴⁵, probably because viruses are more likely to be seen when observing an entire cell, rather than just a portion. However, visualization through a whole cell is less sharp than through thin sections, so intracellular components such as storage products may be more likely to be mistaken for viruses with this method. Nevertheless, the approach has yielded similar estimates of the virus contribution to mortality in aerobic waters (~10–50%), and it has also been used in anaerobic fresh water where percentages have been much higher⁴⁵ (~45–100%).

A simple way to examine the possible impact of viruses on planktonic systems is to artificially increase the number of viruses. This has been done experimentally using natural viral concentrates produced by ultrafiltration of sea water^{46–51}. As little as a 20% increase in the 2–200-nm size fraction of viruses can reduce photosynthetic rates by as much as 50% (refs 46, 48). In a variety

of coastal and open-ocean samples with added viruses, growth inhibition (by about half) of the total bacterial community, thought to be primarily heterotrophs, was observed⁴⁷. Although the concentrates increase the mortality of bacteria, they may also stimulate⁵¹ the growth of the surviving bacteria, as predicted by models^{2,7} and experimental studies^{50,52}. Heat-treating the concentrates removed the inhibitory effect in these studies, which is consistent with a viral effect. This has significant biogeochemical implications (see later).

Viral decay. Although studied for a variety of reasons, historically for sanitary engineering, viral decay is an important parameter in the investigation of the ecological roles of viruses. The term 'decay' is ambiguous, sometimes referring to degradation, sometimes to disappearance of detectability (including removal by adsorption); in most cases, the definition is an operational one set by the mode of study. In cultures, decay is usually measured as the decrease over time in the number of measurable, infectious viral particles (or rather centres of infection), after spiking a sample with culturable viruses. In whole viral communities, decay is measured as the disappearance over time of directly countable viral particles (new virus production must be stopped to measure this⁴²).

In general, measurements of viral decay have led to estimates of viral turnover times (the time it takes for the standing stock to be replaced as it decays) of about an hour to a few days^{42,53–55}. Decay rates tend to follow the overall biological richness of the water, with the fastest decay occurring in rich nearshore waters and the slowest decay in low-nutrient offshore waters.

Studying loss of infectivity and the mechanisms leading to such loss requires cultures of the hosts, so the largely unknown nature of marine microbial diversity has resulted in relatively few marine viruses being studied. A virus undergoing decay might lose its infectivity well before it decays beyond recognition, hence loss of infectivity provides a more sensitive measure of the decay rate⁵⁶.

The most prominent single decay factor identified in several marine studies so far is sunlight-induced damage^{53,57} (mostly from UV-B, but also from other wavelengths). However, susceptibility to damage from full sunlight varies greatly, with some viruses decaying rapidly at a rate of 40–80% per hour^{53,57}, whereas others decay at a slow rate of only 5–10% per hour^{14,55,56}; thus, generalizations are hard to make. Viruses from sunnier regions apparently show more resistance to light damage, an adaptation that one might expect⁵⁵. Because light is attenuated with depth, solar-induced decay is most important near the sea surface and has no effect in the deep sea. But even with light attenuation, the particularly susceptible viruses in the surface mixed layer may be more affected by solar radiation than by any other decay mechanism⁵³.

In the presence of light, host bacteria can often repair light-damaged viral DNA, restoring infectivity⁵⁸. This photoreactivation (or other light-dependent repair mechanism in host cells) has led to the new conclusion that most viruses in the lighted portion of the water column are infective⁵⁸, contrary to the previous common supposition that they are inactive.

Solar radiation is by no means the only major mechanism of viral decay; other mechanisms can be more important in many instances (even in sunlight). Significant decay has been observed⁵⁵ from heat-labile substances that occur in particulate form (that is, bacteria, protists and some aggregates), and also in the form of high molecular mass (>30 kilodaltons) dissolved material such as hydrolytic enzymes. Often the particulate and dissolved material work synergistically with sunlight. Typical overall decay rates are about 3–10% per hour, implying turnover times of about 0.5–1.5 days. The slowest decay rates, typically 0.5–1% per hour, are observed when viruses are placed in heat-treated sea water in the dark.

Early culture experiments⁵⁹ showed that the rate of decline of infectivity in sea water can slow with time, leaving a slowly decaying low level of infectious viruses. A possible explanation is that there are

refuges in sea water that protect the viruses. Alternatively, natural variation among viruses, even within a burst, may lead to some that are assembled perfectly and so are more resistant to decay than their siblings, which may have minor flaws in the protein coat. Therefore, in modelling viral processes, it may be invalid to apply a single decay rate parameter, perhaps even for a single viral type.

Virus production. The quantitative significance of viral activity can be investigated by estimating the rate at which viruses are produced, because this necessarily implies infection of host cells. Assuming that the infection is lytic, burst size (the number of progeny viruses produced from a single ruptured host cell) can be used to estimate the host mortality rate from the viral production rate. Viral production has been estimated in three ways, as described below.

The first method involves comparing viral community decay rates (where production has been halted) to untreated controls in which production continues⁴². Net decay rates reflect turnover and replenishment of viruses; these processes tend to balance out, as virus abundance is relatively steady. The earliest studies used cyanide to inhibit production of new viruses, and found⁴² viral turnover times of about 1–2 h. Some experiments reported surprisingly high viral decay rates, implying losses that were incompatible with independent measurements⁶⁰ of bacterial production, so there is uncertainty in interpreting this approach. But calculations that assume more conservative infectivity decay rates can still lead to estimates of significant viral production. For example, a tenfold excess of viruses over bacteria can be maintained with a lytic burst size of 50 and viral turnover time of one day, if 20% of the bacteria lyse daily.

The second method of determining viral production is more direct, involving the incorporation of radiolabelled thymidine or phosphate tracers in viral nucleic acid⁶¹. This approach has been used in California and Arctic surface waters^{24,54,61}, revealing that viruses account for about 10–40% of the total mortality rate there.

The third method uses fluorescently labelled viruses (concentrated from sea water, stained and then added back). By means of calculations that are analogous to those used for isotope dilution studies, it is possible to calculate simultaneously the production and decay rates of viruses by measuring the change in total and labelled virus counts⁶². This method has been used to measure virus turnover times of about 1–2 days in Southern California nearshore and offshore waters, probably accounting for most of the total bacterial mortality⁶².

Bacterial mortality. It is generally thought^{14,24,45,54,63,64} that viruses are responsible for about 10–50% of the total bacterial mortality in surface waters, and 50–100% in environments that are unfriendly to protists, such as low-oxygen lake waters. Multiple correlation analysis of the abundances of bacteria, viruses and flagellates indicates⁴¹ that virus-induced mortality of bacteria can occasionally prevail over flagellate grazing, especially at high bacterial abundances.

A few studies have directly compared virus-induced mortality with other causes, and balanced total mortality and loss rates with independent estimates of bacterial production. For example, a study⁵⁴ of California coastal waters found evidence that the total mortality balanced production to within 30% (and that viruses were responsible for about 40% of the total mortality). In the Bering and Chukchi (Arctic) Seas, evidence has been found²⁴ that viruses and protists are responsible for a similar amount of bacterial mortality, with protists dominating in some water samples and viruses in others; however, the total mortality estimates typically fell short of balancing production estimates, often by more than 50%. It was also suggested²⁴ that the viral effect is probably larger in eutrophic waters than in oligotrophic waters, although there are still not enough data to make firm conclusions.

Not every study has concluded that viruses have a significant effect on bacterial mortality: relatively little infection was observed⁶⁵ in a Mediterranean saltern pond, and cyanobacterial mortality was hardly affected by viruses in one particular study area⁶⁶ (see later).

Therefore, some groups of organisms and some habitats may display relatively little virus induced mortality; the effect may be seasonal, local or sporadic.

Overall, it seems reasonable to conclude that viruses often, but not always, have a significant effect on bacterial mortality, sometimes even greater than grazing. Several studies indicate that viruses are relatively more important in nutrient-rich waters.

Phytoplankton mortality. This topic is not as extensively studied as infection of heterotrophic bacteria, but its potential impact is great, given that phytoplankton form the base of the marine food web. A few studies have looked at the overall effect of viruses. As mentioned earlier, a relatively modest 20% enrichment of sea water with a native virus concentrate leads to 50% reduction of phytoplankton biomass and primary production, with the effect eliminated by heat treatment^{46,48}. This is strong presumptive evidence that abundant viruses actively infect a significant proportion of the phytoplankton community.

In contrast to these whole-community studies, other reports have focused on particular host organisms. Studies of viruses that infect ecologically important phytoplankton are facilitated because some of these species are available in pure culture, and are also readily recognizable by microscopy (unlike heterotrophs).

Marine phytoplankton are composed of prokaryotes (cyanobacteria and prochlorophytes) and eukaryotes; both of these major types show evidence of viral infection in nature. Highly abundant viruses (often 10^2 – 10^4 per millilitre in nearshore and offshore waters, and sometimes exceeding 10^5 per millilitre) can infect^{66,67} specific strains of marine cyanobacteria (*Synechococcus* sp.). Electron microscopy shows that 1–3% of the native cyanobacteria from a variety of locations contain assembled viruses apparently near the end of the lytic cycle⁶.

However, cyanophages are probably not responsible for a large proportion of the cyanobacterial mortality: about 5–15% of the cyanobacteria in the Gulf of Mexico (Texas coast) are lysed by cyanophages daily⁶⁷, with an even smaller percentage⁶⁶ (up to about 3% per day) of *Synechococcus* lysed in temperate Atlantic waters (near Woods Hole, Massachusetts, and also offshore). As culturable cyanobacteria tend to be resistant to co-occurring viruses, it was concluded⁶⁶ that cyanophages affect the species or strain composition of the community more than they affect the total abundance. This has numerous implications for the importance of viruses in affecting community structure, and perhaps ultimately community function as well (discussed later).

Viruses infecting eukaryotic marine phytoplankton are also found in the marine euphotic zone, and are sometimes quite abundant. Marine viruses^{34,38,68–72} are capable of infecting diatoms, chrysophytes, pyrmnesiophytes, haptophytes, raphidophytes and cryptomonads; probably every kind of phytoplankton can be infected. Several viruses that are pathogenic to the common marine prasinophyte *Micromonas pusilla* (a small flagellate) have been isolated from sea water in several locations^{68,73,74}; their abundance can exceed 10^5 per millilitre in coastal waters and the turnover time is about 1.3 days, with infection leading to 2–10% of the host population lysing per day⁷⁴.

The viruses can be diverse: five genotypes of the *M. pusilla* virus were identified⁷⁵, using sequence analysis of DNA polymerase genes, from one region (offshore Gulf of Mexico). Significant diversity was also found⁷¹ among 14 virus strains isolated and tested against 18 cultured strains of *Heterosigma akashiwo*.

Given the dependence of viral infection on host population density, such infection is expected to be most prominent in algal blooms, where hosts are abundant. Several reports indicate that viral infection may be important in the control of algal blooms. Phytoplankton that are susceptible to viral infection include the chrysophyte *Aureococcus anophagefferans*^{4,70,76} (which causes costly and devastating 'brown tides' in US northeastern embayments), the toxic raphidophyte *Heterosigma akashiwo*⁷¹ (which kills fish in

both aquaculture and natural settings), the common *Phaeocystis pouchetii*⁷⁷ and *Emiliana huxleyi*^{69,78,79} (an organism distributed worldwide that often forms large, dense blooms in mid-latitudes). Up to 50% of all *E. huxleyi* cells in a decaying natural bloom were observed⁷⁹ using TEM to be infected, and viral lysis can account for 25–100% of the net mortality of *E. huxleyi* during the decline of blooms in experimental mesocosms⁶⁹. *E. huxleyi* is one of the planet's major producers of calcite (in the form of coccoliths), causing a large amount of carbon to be transported to sediments⁸⁰; it also generates dimethyl sulphide (DMS), a gas important in regulating climate⁸¹.

Biogeochemical effects

Given that viruses often cause a large proportion of bacterial and even phytoplankton mortality (particularly during blooms), what are their effects on the system as a whole? Lytic infection converts cells into viruses plus cellular debris. This debris is made up of dissolved molecules (monomers, oligomers and polymers) plus colloids and cell fragments⁸², most of which is operationally defined as dissolved organic matter.

What is the likely fate of this material? The most reasonable assumption is that most or all of the lysis products will immediately or eventually become available to bacteria^{2,6,7}. This has recently been confirmed experimentally^{52,62,76}, although a very small proportion of the viruses may be grazed directly by heterotrophic flagellates⁸³, and some cellular debris may resist degradation.

If the cell lysed is a bacterium, then the availability of the lysis products to bacteria represents a semi-closed trophic loop, whereby bacterial biomass is consumed primarily by other bacteria. The loop is fed externally by release of dissolved organic matter from phytoplankton and grazers (Fig. 2). Given that there are respiratory losses and inorganic nutrient regeneration associated with use of dissolved organic substances, this loop has the net effect of oxidizing organic matter and regenerating inorganic nutrients^{2,7} (Fig. 2). However, given that protists would otherwise consume the bacteria¹, this bacterial consumption of matter originating from other bacteria has the interesting property of essentially robbing production from the rest of the food web, and sequestering the biomass and activity into the dissolved and smallest particulate forms².

Similarly, viral lysis of phytoplankton would rob the larger grazers and move material into these small forms. The net effect has been illustrated by a model² showing that, compared to a system with no viruses, an otherwise identical food web with 50% of bacterial mortality from viruses has 27% more bacterial respiration and

production, and 37% less bacterial grazing by protists, which culminates in a 7% reduction in macrozooplankton production.

However, the original steady-state model² assumed that only bacteria are infected and that all the viral matter is consumed by bacteria. A modification of that model, including a small amount of viral infection of phytoplankton (7% loss) and also flagellate grazing of 3% of the virus production, has essentially the same net effect of increasing bacterial production and respiration (by 33%), and reducing protist and animal production (Fig. 3).

Sequestration of materials in viruses, bacteria and dissolved matter may lead to better retention of nutrients in the euphotic zone in virus-infected systems, because more material remains in these small non-sinking forms². This may be particularly important for nutrients that potentially limit productivity (such as N, P and Fe), which are relatively concentrated in bacteria compared to eukaryotes². Reduced viral activity may result in more material in larger organisms, which either sink themselves or as detritus, transporting carbon and inorganic nutrients from the euphotic zone to the deep sea.

Lysis of organisms and release of their cell contents to the water have other potential geochemical effects, because of the chemical and physical properties of the released materials and the location in the water column where the lysis occurs. For example, released polymers contribute to the 'gel' characteristics of sea water that influence many biological and microscopic physical–chemical processes, and may facilitate aggregation and sinking of material from the euphotic zone^{38,49}. Alternatively, viral lysis of microorganisms within sinking aggregates may effectively dissolve the particles, converting some sinking particulate matter into non-sinking dissolved material and colloids at whatever depth the lysis occurs³⁸.

Viruses may also be important in shaping global climate because they induce the release of DMS, a gas that influences cloud nucleation. Experiments show that infection causes the total release of intracellular DMS precursor (DMSP) from the phytoplankter *M. pusilla*⁸⁴, and viruses are also known to infect major DMSP-containing bloom organisms such as *E. huxleyi*⁷⁹ and *P. pouchetii*, causing enhanced release of DMS⁸⁵. Complex interactions within the food web (from microbes to zooplankton) determine whether appreciable DMS can be released to the atmosphere, or whether it is transformed and degraded primarily by microbes or animals in the water⁸⁶. Thus, it is important to understand physical mixing processes and the activities and complex interactions of viruses (both algal and bacterial), protists and zooplankton grazers in order to predict the extent of oceanic DMS release compared with its degradation.

Ecological effects

Diversity regulation. Marine prokaryote diversity is beginning to be understood with the advent of 16S ribosomal RNA sequence-based techniques, which enable analysis of most organisms that are resistant to conventional cultivation⁸⁷. As evident from their effect on algal blooms and cyanobacteria, viruses are in a unique position to influence community species compositions. Theoretical analysis⁸⁸ indicates that the maximum possible number of bacterial populations in a spatially homogeneous environment is equal to the number of unique resources plus the number of unique virus types.

Even if viruses cause only a small proportion of the mortality of a group of organisms, they still can have a profound effect on the relative proportions of different species or strains in the community^{66,89}. Similar ideas have been known from experimental systems for some time⁹. An important consideration is that viral infection is thought to be both density dependent and species specific (or nearly so). Because viruses must diffuse randomly from host to host, rare hosts are less susceptible to the spread of infection than more common ones. Lytic viruses can only increase in abundance when the average time to diffuse from host to host is shorter than the average time they remain infectious. Thus, when a

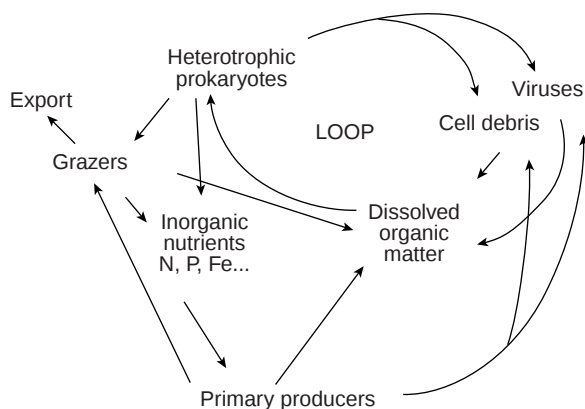


Figure 2 The planktonic viral loop. Schematic diagram of an aquatic food web, emphasizing the semi-closed loop connecting prokaryotes, viruses and dissolved organic matter (curved lines). Note that the loop has the net effect of converting organic matter into dissolved inorganic nutrients. The viruses and cell debris are separated for illustration, but are often defined operationally as dissolved organic matter.

particular species or strain becomes more densely populated, it is more susceptible to infection. This may have direct relevance to solving Hutchinson's 'paradox of plankton'⁹⁰, which asks how so many different kinds of phytoplankton can coexist on only a few potentially limiting resources, when competition theory predicts just one or a few competitive winners. Although several possible explanations may contribute to the answer⁹¹, viral activity probably assists because the competitive dominants become particularly susceptible to infection, whereas rare species are relatively protected¹⁴.

Models⁹² of the potential factors that control the biomass and species composition of microbial systems include growth limitation by organic carbon, inorganic phosphate or nitrogen (inorganic or organic), and cell losses due to grazing by protists and viral lysis. Even when bacterial abundance is assumed to be controlled by protist grazing, the models consistently show that viruses control the steady-state diversity of the bacterial community, whether bacterial growth-rate limitation is by organic or inorganic nutrients. Consistent with this picture are studies^{35,93} of viral genome sizes and hybridization analyses of total viral communities, which show episodic and spatial changes in viral community composition. Thus, both empirical and theoretical analyses indicate that viruses are important in regulating patterns of diversity.

Resistance. The development of resistance to viral infection is well known from non-marine experimental studies⁹, and is an important assumption in models^{88,92} describing the regulation of diversity by viruses. Resistance has been shown⁶⁶ to occur in marine cyanobacteria. The incidence of resistance in marine heterotrophic bacterial populations has not been evaluated directly; doing so would be difficult because most have not yet been grown in culture. However, the high rates of virus production in many marine locations seem to be inconsistent with a community dominated by virus-resistant organisms.

Why is it that viral resistance does not appear to be a dominant factor, even though theory and some field data indicate that it should be? Unfortunately, there are no obvious explanations. One possibility is that resistance is dominant and virus production is not as high as it seems to be, but this would require explanations for the varied observations that are consistent with high virus production. Another possibility is that resistance costs too much physiologically, in that it often occurs through the loss of some important receptor and thus confers a competitive disadvantage⁹. Theory suggests that if the cost is too high, resistant organisms cannot persist in competition with sensitive ones, even when viruses are present⁸⁸. Chemostat experiments⁹⁴ indicate that even with a cost that allows persistence of resistant cells, they may still be rare unless the cost is low or limiting resources are abundant. However, laboratory experiments indicate that low-cost or no-cost resistance is possible⁹, so uncertainty remains.

It seems likely that the explanation lies outside the realm of simplified theoretical or laboratory systems. The real ocean has many species of bacteria and viruses living together, with a variety of alternative mortality mechanisms and low, but variable, nutrient conditions. Perhaps there are advantages to being sensitive to infection. In a typical (oligotrophic) marine environment, bacterial growth may be limited by nitrogen, phosphorus or organic carbon; unsuccessful viral infection (perhaps stopped intracellularly by a restriction enzyme, or with a genetic incompatibility) is then of significant nutritional benefit⁶ to the host organism, because the virus injection of DNA is rich in C, N and P. Even the viral protein coat, left attached to the outside of the cell, is probably digestible by cell-associated exoenzymes⁹⁵. In this situation, one might even imagine bacteria using decoy virus receptors for viral strains that cannot successfully infect them; if an infectious one occasionally gains entry, the strategy might still confer a net benefit to the cell line.

Could a model with so many aborted infections still be consistent

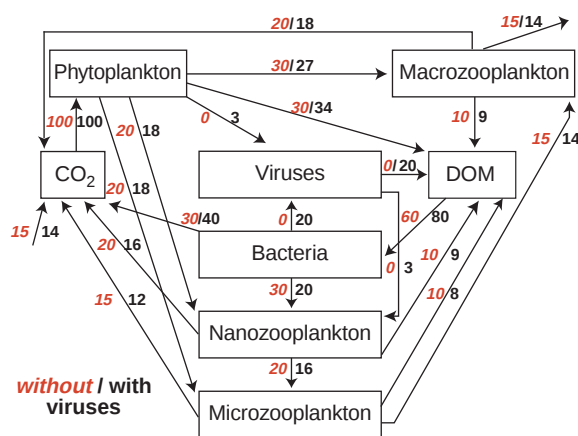


Figure 3 Modelling viral effects on carbon flow. Steady-state model of carbon flow in a hypothetical aquatic microbial food web with and without a virus component that is responsible for 50% of bacterial mortality and 7% of phytoplankton mortality. The numbers represent units of carbon transferred between boxes in a given time unit. The virus compartment includes bacterial cell debris from lysis, and lysed cells are assumed to degrade to dissolved organic matter (DOM). Nanozooplankton are heterotrophic protists 2–20 µm in diameter and microzooplankton are protists or small metazoa ~20–200 µm in diameter (see Box 1 for further descriptions). Note that the viruses lead to an increase in bacterial production and respiration, and a decrease in zooplankton production and respiration. See text and ref. 2 for details.

with high virus production rates? It would seem so given that, in the steady-state, only one virus out of every burst from cell lysis (typically producing 20–50 progeny viruses) successfully infects another host. A large proportion of the remaining 95–98% unsuccessful viruses could end up infecting the 'wrong' hosts, giving these hosts an incentive to continue permitting virus attachment and injection. This is a consistent model, but highly speculative.

Another consideration that argues against resistance relates to the results of system models (Fig. 3) which show that, as a group, the heterotrophic bacteria benefit greatly from viral infection (as infection boosts their production significantly, essentially by taking carbon and energy away from large organisms). Viruses also boost the entire system biomass and production by helping to maintain nutrients in the lighted surface waters. However, these explanations would additionally require some sort of group-selection theory to explain how individuals would benefit from not developing resistance (for example, why not cheat by developing resistance and letting all the other organisms give the group benefits of infection?).

It seems that most explanations of why resistance is not a dominant factor are either unsatisfactory or highly speculative. Whether lack of comprehensive resistance is due to frequent development of new virulent strains, rapid dynamics or patchiness in species compositions, or to a stable coexistence of viruses and their hosts has yet to be determined. As a final note, the extent of resistance may somehow be related to pseudolysogeny (in the sense of a transient immunity to infection induced by infected hosts^{11,12}), but this is not well understood.

Lysogeny. Lysogeny has been suggested as a survival strategy of viruses living at low host density; at the same time, it may confer advantages to the host, including immunity from infection by related viruses and the acquisition of new functions coded by the virus genome (known as conversion⁹). In seawater and lakewater samples^{96,97}, lysogens (bacteria that harbour prophage and can be induced to produce lytic viruses) were found to be common, with variable abundances ranging from undetectable to almost 40% of the total bacteria, and this variability can be seasonal⁹⁸. About 40% of cultured marine bacteria are lysogens⁹⁹.

There may be certain natural events that induce such lysogens to enter the lytic cycle, and reports^{96,100} show that common pollutants such as hydrocarbons, polychlorinated biphenyls (PCBs) and Aroclor can do so. However, growth experiments with native mixed bacterial communities grown in filtered sea water indicate that under typical natural conditions, including exposure to bright sunlight, induction of lysogens is rare, and the vast majority (97% or more) of viruses observed in sea water are probably the result of successive lytic infection¹⁰¹. The same conclusion was reached from a quantitative interpretation of induction experiments^{99,102}. Lyso-genic induction may be occurring at a low level all the time, or may occur sporadically. Although induction may be rare, the widespread existence of lysogeny probably has major implications for genetic exchange and evolution (see below).

Genetic exchange. Probably the most uncertain aspect of marine viruses is their role in genetic exchange among microorganisms, and its effect on short-term adaptation, population genetics and evolution. There can be direct effects such as transduction^{103,104}, in which a virus picks up DNA from one host and transfers it to another (often a lysogen). The overall effect over large scales of space and time would be to homogenize genes among the susceptible host populations. Although the traditional view is that transduction usually operates within a highly restricted host range, one report¹³ indicates that some marine bacterial viruses are capable of nonspecific horizontal gene transfer. Another consideration is that viral lysis causes release of DNA from host organisms, which could be transferred to another organism through natural transformation. Dissolved DNA is readily found in sea water, and it has been reported¹⁹ that viral lysis may be a major source mechanism.

The latter two processes (generalized viral horizontal gene transfer and transformation) would have the effect of mixing genes among a broad variety of species, with wide-ranging effects on adaptation and evolution. Although transfers of these sorts may be extremely rare, the typical bacterial abundances of 10^9 per litre in the euphotic zone and the huge volume of the sea ($3.6 \times 10^7 \text{ km}^3$ in the top 100 m; ref. 105), coupled with generation times on the order of a day, implies that an event with a probability of only 10^{-20} per generation would be occurring about a million times per day. Thus, these process may have major effects on the genetic structure and evolution of the global population of marine bacteria. They should also be considered in the evaluation of the potential spread of genetically engineered microbial genes, or of antibiotic resistance introduced by intensive fish farming.

Looking ahead

New molecular methods that enable the diversity of viruses and their hosts to be studied within their natural habitats will make it easier to unravel the complex web of interactions in marine communities. Once we gain a better understanding of diversity issues and important factors such as host range, we can start to follow, and maybe even predict, the complex relationships between viruses and the rest of the ecosystem. □

Jed Fuhrman is in the Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089-0371, USA.

- Azam, F., Fenchel, T., Gray, J. G., Meyer-Reil, L. A. & Thingstad, T. The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
- Fuhrman, J. A. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 361–383 (Plenum, New York, 1992).
- Fuhrman, J. A., Sletter, T. D., Carlson, C. A. & Proctor, L. M. Dominance of bacterial biomass in the Sargasso Sea and its ecological implications. *Mar. Ecol. Prog. Ser.* **57**, 207–217 (1989).
- Sieburth, J. M., Johnson, P. W. & Hargraves, P. E. Ultrastructure and ecology of *Aureococcus anophagefferens* gen. et sp. nov. (Chrysophyceae): the dominant picoplankton during a bloom in Narragansett Bay, Rhode Island, Summer 1985. *J. Phycol.* **24**, 416–425 (1988).
- Bergh, O., Borsheim, K. Y., Bratbak, G. & Haldal, M. High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468 (1989).
- Proctor, L. M. & Fuhrman, J. A. Viral mortality of marine bacteria and cyanobacteria. *Nature* **343**, 60–62 (1990).
- Bratbak, G., Haldal, M., Norland, S. & Thingstad, T. F. Viruses as partners in spring bloom microbial trophodynamics. *Appl. Environ. Microbiol.* **56**, 1400–1405 (1990).

- Ackermann, H.-W. & DuBow, M. S. *Viruses of Prokaryotes* Vol. 1, *General Properties of Bacteriophages* (CRC, Boca Raton, 1987).
- Lenski, R. E. Dynamics of interactions between bacteria and virulent bacteriophage. *Adv. Microb. Ecol.* **10**, 1–44 (1988).
- Ripp, S. & Miller, R. V. The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology* **143**, 2065–2070 (1997).
- Moebus, K. Marine bacteriophage reproduction under nutrient-limited growth of host bacteria. 2. Investigations with phage-host system [H3:H3V1]. *Mar. Ecol. Prog. Ser.* **144**, 13–22 (1996).
- Moebus, K. Investigations of the marine lysogenic bacterium H24. II. Development of pseudolysogeny in nutrient rich broth. *Mar. Ecol. Prog. Ser.* **148**, 229–240 (1997).
- Chiura, H. X. Generalized gene transfer by virus-like particles from marine bacteria. *Aquat. Microb. Ecol.* **13**, 75–83 (1997).
- Fuhrman, J. A. & Suttle, C. A. Viruses in marine planktonic systems. *Oceanography* **6**, 51–63 (1993).
- Sharp, G. D. Enumeration of virus particles by electron microscopy. *Proc. Soc. Exp. Biol. Med.* **70**, 54–59 (1949).
- Wommack, K. E., Hill, R. T., Kessel, M., Russek-Cohen, E. & Colwell, R. R. Distribution of viruses in the Chesapeake Bay. *Appl. Environ. Microbiol.* **58**, 2965–2970 (1992).
- Borsheim, K. Y. Native marine bacteriophages. *FEMS Microbiol. Ecol.* **102**, 141–159 (1993).
- Cochlan, W. P., Wikner, J., Steward, G. F., Smith, D. C. & Azam, F. Spatial distribution of viruses, bacteria and chlorophyll *a* in neritic, oceanic and estuarine environments. *Mar. Ecol. Prog. Ser.* **92**, 77–87 (1993).
- Paul, J. H., Rose, J. B., Jiang, S. C., Kellogg, C. A. & Dickson, L. Distribution of viral abundance in the reef environment of Key Largo, Florida. *Appl. Environ. Microbiol.* **59**, 718–724 (1993).
- Boehme, J. et al. Viruses, bacterioplankton, and phytoplankton in the southeastern Gulf of Mexico: distribution and contribution to oceanic DNA pools. *Mar. Ecol. Prog. Ser.* **97**, 1–10 (1993).
- Maranger, R., Bird, D. F. & Juniper, S. K. Viral and bacterial dynamics in arctic sea ice during the spring algal bloom near Resolute, NWT, Canada. *Mar. Ecol. Prog. Ser.* **111**, 121–127 (1994).
- Hara, S., Koike, I., Terauchi, K., Kamiya, H. & Tanoue, E. Abundance of viruses in deep oceanic waters. *Mar. Ecol. Prog. Ser.* **145**, 269–277 (1996).
- Maranger, R. & Bird, D. E. High concentrations of viruses in the sediments of Lac Gilbert, Quebec. *Microb. Ecol.* **31**, 141–151 (1996).
- Steward, G. F., Smith, D. C. & Azam, F. Abundance and production of bacteria and viruses in the Bering and Chukchi Sea. *Mar. Ecol. Prog. Ser.* **131**, 287–300 (1996).
- Noble, R. T. & Fuhrman, J. A. Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. *Aquat. Microb. Ecol.* **14**, 113–118 (1998).
- Hennes, K. P. & Simon, M. Significance of bacteriophages for controlling bacterioplankton growth in a mesotrophic lake. *Appl. Environ. Microbiol.* **61**, 333–340 (1995).
- Bratbak, G., Haldal, M., Thingstad, T. F. & Tuomi, P. Dynamics of virus abundance in coastal seawater. *FEMS Microb. Ecol.* **19**, 263–269 (1996).
- Fuhrman, J. A. & Ouverney, C. C. Marine microbial diversity studied via 16S rRNA sequences: cloning results from coastal waters and counting of native archaea with fluorescent single cell probes. *Aquat. Ecol.* **32**, 3–15 (1998).
- Hara, S., Terauchi, K. & Koike, I. Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy. *Appl. Environ. Microbiol.* **57**, 2731–2734 (1991).
- Hennes, K. P. & Suttle, C. A. Direct counts of viruses in natural waters and laboratory cultures by epifluorescence microscopy. *Limnol. Oceanogr.* **40**, 1050–1055 (1995).
- Weinbauer, M. G. & Suttle, C. A. Comparison of epifluorescence and transmission electron microscopy for counting viruses in natural marine waters. *Aquat. Microb. Ecol.* **13**, 225–232 (1997).
- Xenopoulos, M. A. & Bird, D. F. Microwave enhanced staining for counting viruses by epifluorescence microscopy. *Limnol. Oceanogr.* **42**, 1648–1650 (1997).
- Marie, D., Brussaard, C. P. D., Thyrhaug, R., Bratbak, G. & Vaulot, D. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl. Environ. Microbiol.* **65**, 45–52 (1999).
- Proctor, L. M. Advances in the study of marine viruses. *Microsc. Res. Tech.* **37**, 136–161 (1997).
- Wommack, K. E., Ravel, J., Hill, R. T., Chun, J. S. & Colwell, R. R. Population dynamics of Chesapeake Bay viroplankton: total-community analysis by pulsed-field gel electrophoresis. *Appl. Environ. Microbiol.* **65**, 231–240 (1999).
- Steward, G. F. & Azam, F. in *Microbial Biosystems: New Frontiers. Proc. 8th Int. Symp. Microb. Ecol.* (eds Bell, C. R., Brylinsky, M. & Johnson-Green, P.) (Atlantic Canada Society for Microbial Ecology, Halifax, Canada, 1999).
- Murray, A. G. & Jackson, G. A. Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. *Mar. Ecol. Prog. Ser.* **89**, 103–116 (1992).
- Proctor, L. M. & Fuhrman, J. A. Roles of viral infection in organic particle flux. *Mar. Ecol. Prog. Ser.* **69**, 133–142 (1991).
- Mathias, C. B., Kirschner, A. K. T. & Velimirov, B. Seasonal variations of virus abundance and viral control of the bacterial production in a backwater system of the Danube River. *Appl. Environ. Microbiol.* **61**, 3734–3740 (1995).
- Proctor, L. M., Okubo, A. & Fuhrman, J. A. Calibrating estimates of phage induced mortality in marine bacteria: ultrastructural studies of marine bacteriophage development from one-step growth experiments. *Microb. Ecol.* **25**, 161–182 (1993).
- Weinbauer, M. G. & Peduzzi, P. Significance of viruses versus heterotrophic nanoflagellates for controlling bacterial abundance in the northern Adriatic Sea. *J. Plank. Res.* **17**, 1851–1856 (1995).
- Heldal, M. & Bratbak, G. Production and decay of viruses in aquatic environments. *Mar. Ecol. Prog. Ser.* **72**, 205–212 (1991).
- Weinbauer, M. G., Fuks, D. & Peduzzi, P. Distribution of viruses and dissolved DNA along a coastal trophic gradient in the Northern Adriatic Sea. *Appl. Environ. Microbiol.* **59**, 4074–4082 (1993).
- Weinbauer, M. G. & Peduzzi, P. Frequency, size, and distribution of bacteriophages in different marine bacterial morphotypes. *Mar. Ecol. Prog. Ser.* **108**, 11–20 (1994).
- Weinbauer, M. G. & Hofle, M. G. Significance of viral lysis and flagellate grazing as factors controlling bacterioplankton production in a eutrophic lake. *Appl. Environ. Microbiol.* **64**, 431–438 (1998).
- Suttle, C. A., Chan, A. M. & Cottrell, M. T. Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* **387**, 467–469 (1990).
- Proctor, L. M. & Fuhrman, J. A. Mortality of marine bacteria in response to enrichments of the virus size fraction from seawater. *Mar. Ecol. Prog. Ser.* **87**, 283–293 (1992).
- Suttle, C. A. Inhibition of photosynthesis in phytoplankton by the submicron size fraction concentrated from seawater. *Mar. Ecol. Prog. Ser.* **87**, 105–112 (1992).
- Peduzzi, P. & Weinbauer, M. G. Effect of concentrating the virus-rich 2–200 nm size fraction of seawater on the formation of algal flocs (marine snow). *Limnol. Oceanogr.* **38**, 1562–1565 (1993).
- Weinbauer, M. G. & Peduzzi, P. Effect of virus-rich high molecular weight concentrates of seawater on the dynamics of dissolved amino acids and carbohydrates. *Mar. Ecol. Prog. Ser.* **127**, 245–253 (1995).
- Noble, R. T., Middelboe, M. & Fuhrman, J. A. The effects of viral enrichment on the mortality and growth of heterotrophic bacterioplankton. *Aquat. Microb. Ecol.* (in the press).

52. Middelboe, M., Jørgensen, N. O. G. & Kroer, N. Effects of viruses on nutrient turnover and growth efficiency of non-infected marine bacterioplankton. *Appl. Environ. Microbiol.* **62**, 1991–1997 (1996).
53. Suttle, C. A. & Chen, F. Mechanisms and rates of decay of marine viruses in seawater. *Appl. Environ. Microbiol.* **58**, 3721–3729 (1992).
54. Fuhrman, J. A. & Noble, R. T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol. Oceanogr.* **40**, 1236–1242 (1995).
55. Noble, R. T. & Fuhrman, J. A. Virus decay and its causes in coastal waters. *Appl. Environ. Microbiol.* **63**, 77–83 (1997).
56. Wommack, K. E., Hill, R. T., Muller, T. A. & Colwell, R. R. Effects of sunlight on bacteriophage viability and structure. *Appl. Environ. Microbiol.* **62**, 1336–1341 (1996).
57. Wilhelm, S. W., Weinbauer, M. G., Suttle, C. A. & Jeffrey, W. H. The role of sunlight in the removal and repair of viruses in the sea. *Limnol. Oceanogr.* **43**, 586–592 (1998).
58. Wilhelm, S. W., Weinbauer, M. G., Suttle, C. A., Pledger, R. J. & Mitchell, D. L. Measurements of DNA damage and photoreactivation imply that most viruses in marine surface waters are infective. *Aquat. Microb. Ecol.* **14**, 215–222 (1998).
59. Zachary, A. An ecological study of bacteriophages of *Vibrio natriegens*. *Can. J. Microbiol.* **24**, 321–324 (1978).
60. Bratbak, G., Haldal, M., Thingstad, T. F., Riemann, B. & Haslund, O. H. Incorporation of viruses into the budget of microbial C-transfer. A first approach. *Mar. Ecol. Prog. Ser.* **83**, 273–280 (1992).
61. Steward, G. F., Wikner, J., Cochlan, W. P., Smith, D. C. & Azam, F. Estimation of virus production in the sea: II. Field results. *Mar. Microb. Food Webs* **6**, 79–90 (1992).
62. Noble, R. T. The Fates of Viruses in the Marine Environment Thesis, Univ. Southern California (1998).
63. Suttle, C. A. The significance of viruses to mortality in aquatic microbial communities. *Microb. Ecol.* **28**, 237–243 (1994).
64. Weinbauer, M. G., Fuks, D., Puskaric, S. & Peduzzi, P. Diel, seasonal, and depth-related variability of viruses and dissolved DNA in the Northern Adriatic Sea. *Microb. Ecol.* **30**, 25–41 (1995).
65. Guixa-Boixareu, N., Calderon-Paz, J. I., Haldal, M., Bratbak, G. & Pedros-Alio, C. Viral lysis and bacteriophage as prokaryotic loss factors along a salinity gradient. *Aquat. Microb. Ecol.* **11**, 215–227 (1996).
66. Waterbury, J. B. & Valois, F. W. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl. Environ. Microbiol.* **59**, 3393–3399 (1993).
67. Suttle, C. A. & Chan, A. M. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl. Environ. Microbiol.* **60**, 3167–3174 (1994).
68. Suttle, C. A., Chan, A. M. & Cottrell, M. T. Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Appl. Environ. Microbiol.* **57**, 721–726 (1991).
69. Bratbak, G., Egge, J. K. & Haldal, M. Viral mortality of the marine alga *Emiliana huxleyi* (Haptophyceae) and termination of algal blooms. *Mar. Ecol. Prog. Ser.* **93**, 39–48 (1993).
70. Milligan, K. L. D. & Cosper, E. M. Isolation of virus capable of lysing the brown tide microalga, *Aureococcus anophagefferens*. *Science* **266**, 805–807 (1994).
71. Nagasaki, K. & Yamaguchi, M. Intra-species host specificity of HaV (*Heterosigma akashiwo* virus) clones. *Aquat. Microb. Ecol.* **14**, 109–112 (1998).
72. Suttle, C. A. & Chan, A. M. Viruses infecting the marine Prymnesiophyte *Chrysochromulina* spp.: isolation, preliminary characterization and natural abundance. *Mar. Ecol. Prog. Ser.* **118**, 275–282 (1995).
73. Mayer, J. A. & Taylor, F. J. R. A virus which lyses the marine nanoflagellate *Micromonas pusilla*. *Nature* **281**, 299–301 (1979).
74. Cottrell, M. T. & Suttle, C. A. Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate *Micromonas pusilla*. *Limnol. Oceanogr.* **40**, 730–739 (1995).
75. Chen, F., Suttle, C. A. & Short, S. M. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Appl. Environ. Microbiol.* **62**, 2869–2874 (1996).
76. Gobler, C. J., Hutchins, D. A., Fisher, N. S., Cosper, E. M. & SanudoWilhelmy, S. A. Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine chrysophyte. *Limnol. Oceanogr.* **42**, 1492–1504 (1997).
77. Jacobsen, A., Bratbak, G. & Haldal, M. Isolation and characterization of a virus infecting *Phaeocystis pouchetii* (Prymnesiophyceae). *J. Phycol.* **32**, 923–927 (1996).
78. Bratbak, G. et al. Viral activity in relation to *Emiliana huxleyi* blooms—a mechanism of DMSP release. *Mar. Ecol. Prog. Ser.* **128**, 133–142 (1995).
79. Brussaard, C. P. D., Kempers, R. S., Kop, A. J., Riegman, R. & Haldal, M. Virus-like particles in a summer bloom of *Emiliana huxleyi* in the North Sea. *Aquat. Microb. Ecol.* **10**, 105–113 (1996).
80. Dymond, J. & Lyle, M. Flux comparisons between sediments and sediment traps in the Eastern Tropical Pacific: implications for atmospheric CO₂ variations during the Pleistocene. *Limnol. Oceanogr.* **30**, 699–712 (1995).
81. Malin, G., Turner, S. M., Liss, P. S. & Aiken, Sulfur—the plankton climate connection. *J. Phycol.* **28**, 590–597 (1992).
82. Shibata, A., Kogure, K., Koike, I. & Ohwada, K. Formation of submicron colloidal particles from marine bacteria by viral infection. *Mar. Ecol. Prog. Ser.* **155**, 303–307 (1997).
83. Gonzalez, J. M. & Suttle, C. A. Grazing by marine nanoflagellates on viruses and virus-sized particles—ingestion and digestion. *Mar. Ecol. Prog. Ser.* **94**, 1–10 (1993).
84. Hill, R. W., White, B. A., Cottrell, M. T. & Dacey, J. W. H. Virus-mediated total release of dimethylsulfoniopropionate from marine phytoplankton: a potential climate process. *Aquat. Microb. Ecol.* **14**, 1–6 (1998).
85. Malin, G., Wilson, W. H., Bratbak, G., Liss, P. S. & Mann, N. H. Elevated production of dimethylsulfide resulting from viral infection of cultures of *Phaeocystis pouchetii*. *Limnol. Oceanogr.* **43**, 1389–1393 (1998).
86. Kiene, R. P. & Bates, T. S. Biological removal of dimethyl sulphide from sea water. *Nature* **345**, 702–705 (1990).
87. Fuhrman, J. A. in *Manual of Environmental Microbiology* (eds Hurst, C. J., Knudsen, C. R., McInerney, M. J., Stetzenbach, L. D. & Walter, M. V.) 1–894 (American Society for Microbiology, Washington DC, 1997).
88. Levin, B. R., Steward, F. M. & Chao, L. Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. *Am. Nat.* **111**, 3 (1977).
89. Hennes, K. P., Suttle, C. A. & Chan, A. M. Fluorescently labeled virus probes show that natural virus populations can control the structure of marine microbial communities. *Appl. Environ. Microbiol.* **61**, 3623–3627 (1995).
90. Hutchinson, G. E. The paradox of the plankton. *Am. Nat.* **45**, 137–145 (1961).
91. Siegel, D. A. Resource competition in a discrete environment: why are plankton distributions paradoxical? *Limnol. Oceanogr.* **43**, 1133–1146 (1998).
92. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**, 19–27 (1997).
93. Wommack, K. E., Ravel, J., Hill, R. T. & Colwell, R. R. Hybridization analysis of Chesapeake Bay viroplankton. *Appl. Environ. Microbiol.* **65**, 241–250 (1999).
94. Bohannon, B. J. M. & Lenski, R. E. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. *Ecology* **78**, 2303–2315 (1997).
95. Hollibaugh, J. T. & Azam, F. Microbial degradation of dissolved proteins in the seawater. *Limnol. Oceanogr.* **28**, 1104–1116 (1983).
96. Jiang, S. C. & Paul, J. H. Occurrence by lysogenic bacteria in marine microbial communities as determined by prophage induction. *Mar. Ecol. Prog. Ser.* **142**, 27–38 (1996).
97. Tapper, M. A. & Hicks, R. E. Temperate viruses and lysogeny in Lake Superior bacterioplankton. *Limnol. Oceanogr.* **43**, 95–103 (1998).
98. Cochran, P. K. & Paul, J. H. Seasonal abundance of lysogenic bacteria in a subtropical estuary. *Appl. Environ. Microbiol.* **64**, 2308–2312 (1998).
99. Jiang, S. C. & Paul, J. H. Significance of lysogeny in the marine environment—studies with isolates and a model of lysogenic phage production. *Microb. Ecol.* **35**, 235–243 (1998).
100. Cochran, P. K., Kellogg, C. A. & Paul, J. H. Prophage induction of indigenous marine lysogenic bacteria by environmental pollutants. *Mar. Ecol. Prog. Ser.* **164**, 124–133 (1998).
101. Wilcox, R. M. & Fuhrman, J. A. Bacterial viruses in coastal seawater: lytic rather than lysogenic production. *Mar. Ecol. Prog. Ser.* **114**, 35–45 (1994).
102. Weinbauer, M. G. & Suttle, C. A. Potential significance of lysogeny to bacteriophage production and bacterial mortality in coastal waters of the Gulf-of-Mexico. *Appl. Environ. Microbiol.* **62**, 4374–4380 (1996).
103. Ripp, S., Ogunseit, O. A. & Miller, R. V. Transduction of a freshwater microbial community by a new *Pseudomonas aeruginosa* generalized transducing phage, Utl. *Mol. Ecol.* **3**, 121–126 (1994).
104. Jiang, S. C. & Paul, J. H. Gene transfer by transduction in the marine environment. *Appl. Environ. Microbiol.* **64**, 2780–2787 (1998).
105. Sverdrup, H. U., Johnson, M. W. & Fleming, R. H. *The Oceans* (Prentice Hall, Englewood Cliffs, 1942).

Acknowledgements. I thank several individuals who have discussed the broad array of virus-related processes with me over the years and shared unpublished data, including F. Azam, L. Proctor, G. Bratbak, J. Paul, C. Suttle, R. Noble, F. Thingstad, G. Steward and R. Kiene. D. Comeau, B. Bohannon and D. Bird provided useful comments on the manuscript. This work was supported by the NSF and a USC Sea Grant.

Correspondence and requests for materials should be addressed to J.F. (e-mail: fuhrman@usc.edu).

Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1

Hongxu Qin*, Srinivasa M. Srinivasula†‡, Geng Wu*†, Teresa Fernandes-Alnemri‡, Emad S. Alnemri‡ & Yigong Shi§

* Department of Chemistry and § Department of Molecular Biology, Princeton University, Washington Road, Princeton, New Jersey 08544, USA

‡ Department of Microbiology and Immunology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

† These authors contributed equally to this work

Caspase-9-mediated apoptosis (programmed cell death) plays a central role in the development and homeostasis of all multicellular organisms. Mature caspase-9 is derived from its procaspase precursor as a result of recruitment by the activating factor Apaf-1. The crystal structures of the caspase-recruitment domain of Apaf-1 by itself and in complex with the prodomain of procaspase-9 have been determined at 1.6 and 2.5 Å resolution, respectively. These structures and other evidence reveal that each molecule of Apaf-1 interacts with a molecule of procaspase-9 through two highly charged and complementary surfaces formed by non-conserved residues; these surfaces determine recognition specificity through networks of intermolecular hydrogen bonds and van der Waals interactions. Mutation of the important interface residues in procaspase-9 or Apaf-1 prevents or reduces activation of procaspase-9 in a cell-free system. Wild-type, but not mutant, prodomains of caspase-9 completely inhibit catalytic processing of procaspase-9. Furthermore, analysis of homologues from *Caenorhabditis elegans* indicates that recruitment of CED-3 by CED-4 is probably mediated by the same set of conserved structural motifs, with a corresponding change in the specificity-determining residues.

Apoptosis, the prevalent form of programmed cell death, is central to the development and homeostasis of all metazoans^{1–4}. Alterations in apoptotic pathways have been implicated in many human diseases, including cancer and autoimmune disease, and in developmental and possibly neurodegenerative disorders^{5,6}. The mechanism of apoptosis is remarkably conserved across species, executed with a cascade of sequential activation of initiator and effector caspases^{7,8}.

Caspases are a family of cysteine proteases with aspartate substrate specificity, and are produced in cells as inactive zymogens⁷. An effector caspase is activated by an initiator caspase by proteolytic cleavage at specific internal aspartate residues to separate the large and small subunits of the mature caspase⁷. Once activated, the effector caspases can cleave a broad range of cellular targets and ultimately cause cell death. Activation of the initiator caspases is thus an irreversible trigger of apoptosis.

In *C. elegans*, CED-3 is the apoptotic initiator caspase^{3,4}. Activation of CED-3 involves direct recognition of its characteristic prodomain by the caspase-recruitment domain (CARD) of another protein, CED-4. The pro-apoptotic activity of CED-4 is in turn regulated by binding of the anti-apoptotic protein CED-9. During *C. elegans* development, the interaction between CED-3, CED-4 and CED-9 results in the death of 131 cells at precise locations and times.

The mammalian homologues of CED-3 and CED-4 are caspase-9 and the apoptotic protease-activating factor 1, or Apaf-1 (refs 9, 10), respectively. These play a dominant role in the regulation of programmed cell death in mammalian development and in oncogene- and p53-dependent apoptosis^{11–14}. Caspase-9 activation may precede the activation of all other cell-death-related caspases in the mitochondrial pathways of apoptosis¹⁵. An apoptotic stimulus releases cytochrome *c* from the mitochondrial intermembrane space into the cytoplasm, where it associates with Apaf-1 in the presence of dATP or ATP and induces the oligomerization of Apaf-1 (refs 16–18). The oligomeric Apaf-1 complex recognizes the inactive procaspase-9 and forms the so-called apoptosome which triggers autocatalytic processing of procaspase-9 (refs 16–20).

Recognition of procaspase-9 by Apaf-1, primarily through the CARD-prodomain interaction, is essential for formation of the apoptosome and activation of the initiator caspase¹⁰. The 97-residue CARD domain of Apaf-1 shares ~20% sequence identity with the prodomain of procaspase-9, which also belongs to the CARD family of apoptotic signaling motifs²¹. The CARD-homology domain is present in several other caspase-activating proteins, such as CED-4 and RAIDD/CRADD^{22,23}, and in other initiator caspases, such as CED-3 and caspase-2/ICH-1 (ref. 24).

The solution structure of the RAIDD CARD domain determined by NMR reveals a tightly packed globular fold consisting of six antiparallel α -helices²⁵. The surface of RAIDD CARD contains a basic and an acidic patch on opposite sides²⁵. Modelling studies indicated that the basic surface of Apaf-1 CARD could interact with the acidic surface of the procaspase-9 prodomain²⁵.

We now report the crystal structure of Apaf-1 CARD determined at 1.6 Å resolution, which reveals significant structural differences from the RAIDD CARD. Our mutational analysis of Apaf-1 CARD and the procaspase-9 prodomain identifies interaction surfaces on Apaf-1 and procaspase-9 that are unlike those predicted previously²⁵. We also describe the crystal structure of Apaf-1 CARD in complex with the procaspase-9 prodomain at 2.5 Å resolution, which complements our mutagenesis results.

Structure of Apaf-1 CARD

The crystal structure of Apaf-1 CARD, determined at 1.6 Å resolution (Table 1), adopts a compact globular fold, with seven α -helices tightly packed around a central hydrophobic core (Fig. 1a). Of the seven helices, H2, H3, H4 and H5 form an antiparallel four-helix bundle, with helices H1a/H1b and H6 capping the exposed hydrophobic surface of H4 and H5.

The surface of Apaf-1 CARD contains two highly charged surface patches located on two adjacent sides of the molecule (Fig. 1b). On one side, a positively charged patch consists of five basic residues, Arg 13, Lys 42, Lys 58, Lys 62 and Lys 63, which are mainly located on helix H4 and the turn between H1a and H1b. On the adjacent

side, three acidic residues on helices H2 and H3, Asp 27, Glu 40 and Glu 41, constitute a stripe of negatively charged surface, which is capped laterally by two serine residues on H2, Ser 23 and Ser 31 (Fig. 1a). The basic residues create a concave surface, whereas the acidic ones form a convex surface (Fig. 1b).

Although the amino-acid sequence of Apaf-1 CARD only shares 13% identity with RAIDD CARD, the overall X-ray structure of Apaf-1 CARD is similar to the structure of RAIDD CARD determined by multidimensional NMR²⁵. There are some significant structural differences, however, between the two, resulting in over 3.2 Å root-mean-square deviation (r.m.s.d.) for 90 aligned C α atoms (Fig. 1c). For example, the H1 helix of RAIDD CARD is only slightly bent (15°) and retains its integrity as a single α -helix through maintenance of all intrahelical hydrogen bonds²⁵. In contrast, the H1 helix of Apaf-1 CARD is severely bent (54°), and is stretched between residues 11 and 14, resulting in the disruption of hydrogen bonds between the amide nitrogen atoms of residues 12 and 13, and the carbonyl oxygen atoms of residues 16 and 17, respectively. Consequently, the H1 helix in Apaf-1 is now composed of two short helices H1a and H1b, with a two-residue intervening linker (Fig. 1). Because of the bending, residues located at the C-terminal end of helix H1 in Apaf-1 are 4–5 Å closer to the hydrophobic core than are the corresponding residues in RAIDD CARD. This helical shift is propagated to the rest of the molecule, altering the positions and boundaries of all other helices and surface loops (Fig. 1c).

Apaf-1/procaspase-9 interaction

On the basis of sequence conservation, the prodomain of procaspase-9 should have similar structure and surface features. Because of the prominent locations of the charged surface residues (Fig. 1a), we reasoned that they might be involved in the binding between Apaf-1 and procaspase-9, with the basic surface from one molecule binding to the acidic surface of the other through electrostatic interaction. Such an interaction should result in a 1:1 binding stoichiometry between Apaf-1 and procaspase-9: indeed, we found that one molecule of Apaf-1 CARD bound one molecule of the procaspase-9 prodomain *in vitro* (data not shown).

To map the interacting interface between Apaf-1 and procaspase-9, we generated twenty point mutations on the surfaces of Apaf-1 CARD and the procaspase-9 prodomain (Table 2); the mutant residues were mainly on the two highly charged surfaces on each

protein domain. The mutant proteins were all purified to homogeneity and their binding capacity assayed *in vitro* (Table 2).

On the Apaf-1 CARD domain, two mutations, Asp27Ala on helix H2 and Glu40Ala on helix H3, eliminated interaction with the wild-type procaspase-9 prodomain (Table 2). A third mutation, Ser31Ala, weakened but failed to abolish the interaction. Together, these three mutations define the core of the observed acidic surface on the structure of Apaf-1 CARD (Fig. 1a). In contrast, six mutations affecting other residues, including four (Arg 13, Lys 42, Lys 58, Lys 62) of the five basic residues that constitute the positively charged face, had no detectable effect on complex formation (Table 2).

On the procaspase-9 prodomain, two mutations, Arg13Ala at the predicted hinge region between H1a and H1b helices and Arg56Ala on helix H4, prevented interaction with wild-type Apaf-1 CARD (Table 2). Two additional mutations, Arg11Ala at the end of helix H1 and Arg52Ala on helix H4, also significantly reduced the interaction (Table 2). In contrast, seven mutations affecting other residues, including the three acidic residues Glu 18, Asp 23 and Asp 27, which make up the predicted negatively charged face, had no detectable effect on complex formation (Table 2).

These mutagenesis results have identified the structural motifs in Apaf-1 and procaspase-9 necessary for their mutual recognition. The acidic and convex surface of Apaf-1 CARD formed by helices H2 and H3 is directly involved in the recognition and interaction with the predicted basic and concave surface of the procaspase-9 prodomain formed by helices H1a/H1b and H4. This result does not conform with an earlier model of interaction between Apaf-1 and caspase-9 (ref. 25).

Structure of the complex

To pinpoint the molecular basis of procaspase-9 recruitment by Apaf-1, we determined the 2.5 Å crystal structure of Apaf-1 CARD (residues 1–97) in complex with the procaspase-9 prodomain (residues 1–112) (Table 3). Both Apaf-1 CARD and the procaspase-9 prodomain adopt a similar globular fold, each consisting of seven antiparallel α -helices (Fig. 2). The short loops connecting the helices are rigid and well ordered, as defined by good electron density in the crystals and low temperature factors in the final atomic model.

These two domains interact through two complementary surfaces. The slightly concave surface of procaspase-9 formed by the

Table 1 Crystallographic analysis of Apaf-1 CARD

	Reflections		Overall (outer shell)			MIR analysis (20–2.1 Å)			
	Resolution (Å)	Measured	Unique	Data coverage (%)	R_{sym} (%)	Heavy-atom sites	All (outer shell) Isomorphous difference (%)	Acentric/centric Phasing power	Cullis R factor
Native 1 (–150 °C)	1.6	92,435	11,042	96.8 (88.8)	5.7 (12.0)				
Native 2 (20 °C)	2.1	35,362	5,010	93.3 (70.9)	5.8 (9.5)				
HgCl ₂	2.1	8,765	4,854	90.2 (73.1)	3.9 (5.7)	1	21.2 (24.7)	3.91/2.46	0.40/0.49
Se-Met	2.1	14,466	4,900	90.8 (63.3)	7.6 (14.5)	5	14.6 (14.4)	3.24/2.45	0.43/0.44
K ₂ AuCl ₂	2.1	3,423	2,540	47.4 (37.8)	4.5 (8.0)	2	8.5 (10.8)	1.66/1.21	0.70/0.68
K ₂ PtCl ₄	2.1	11,264	4,919	91.4 (84.8)	4.6 (6.8)	3	17.1 (12.1)	1.22/0.91	0.77/0.71
Se-Met/HgCl ₂	2.1	25,164	4,977	92.1 (64.8)	7.9 (13.4)	5/1	31.2 (33.5)	4.47/3.12	0.34/0.38
Se-Met/K ₂ PtCl ₄	2.1	19,996	4,978	92.4 (68.5)	8.8 (17.9)	5/1	22.7 (24.8)	2.77/2.28	0.49/0.45
Thimerosal 1	2.1	9,752	4,881	90.7 (68.5)	4.1 (5.3)	1	11.0 (10.1)	1.77/0.58	0.90/0.89
Thimerosal 2	2.1	9,940	4,567	84.9 (61.6)	4.3 (4.8)	1	8.8 (10.1)	1.96/1.28	0.65/0.66
Mean figure of merit (20.0–2.1 Å): 0.855									
Refinement	Resolution	Reflections	Atoms modelled	$R_{\text{working}}/R_{\text{free}}$ (%)		r.m.s. deviation			
	range	(all)	(total, water)			Bonds (Å)	Angles (deg)	B factor (Å ²)	
	8.0–1.6	10,948	879, 136	19.8/23.6		0.006	1.275	1.768	

Se: Selenium-Met-derivatized Apaf-1 CARD. $R_{\text{sym}} = \sum_i |I_i - \langle I_i \rangle| / \sum_i I_i$, where I_i is the mean intensity of the i observations of symmetry-related reflections of h . Isomorphous difference = $\sum_i |F_{\text{PH}} - F_{\text{P}}| / \sum_i |F_{\text{PH}}|$, where F_{P} and F_{PH} are the native and derivative structure factor amplitudes, respectively. Phasing power = $[(F_{\text{H(calc)}}/F_{\text{PH(obs)}}) - (F_{\text{H(calc)}}/F_{\text{P}})]^{1/2}$, where $F_{\text{PH(obs)}}$ and $F_{\text{H(calc)}}$ are the observed and calculated heavy-atom structure factor. Figure of merit = $(\sum_i P(\alpha) \exp(i\alpha) / \sum_i P(\alpha))$, where $P(\alpha)$ is the probability distribution for the phase α . $R = \sum_i |F_{\text{obs}} - F_{\text{calc}}| / \sum_i F_{\text{obs}}$, where $F_{\text{obs}} = F_{\text{P}}$, and F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections). Root-mean-square (r.m.s.) of bond lengths and angles are the deviations from ideal values; the r.m.s. deviation in B factors is calculated between bonded atoms.

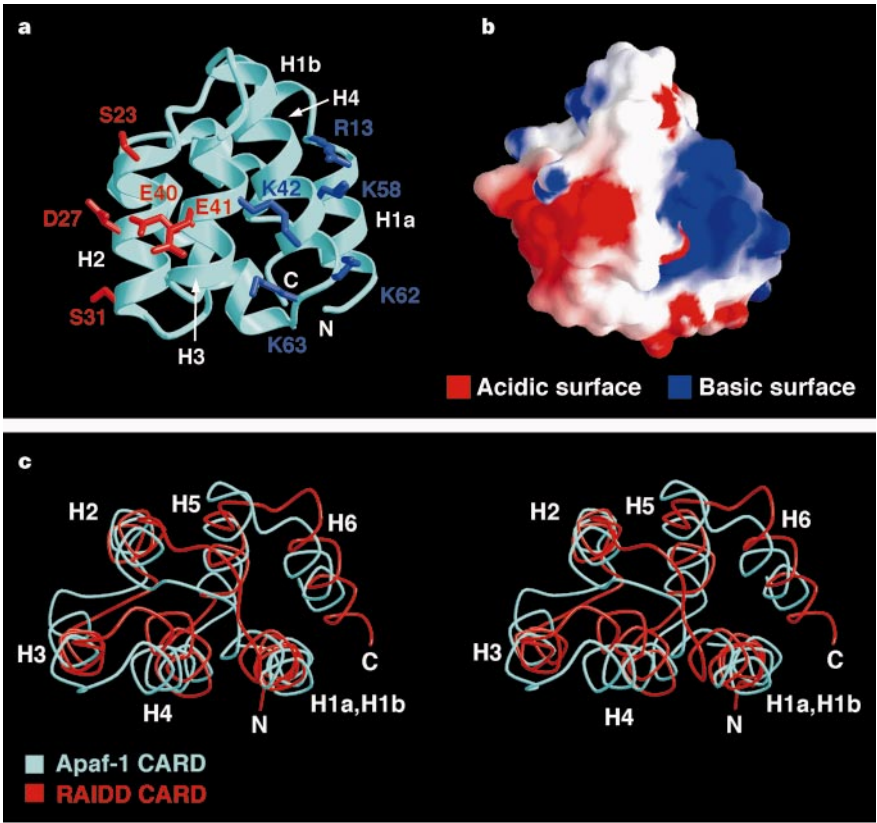


Figure 1 Structure of Apaf-1 CARD domain. **a**, In the structure, the acidic residues on helices H2/H3 and the basic residues on helices H1a/H1b/H4 are represented by red and blue sticks, respectively. **b**, Electrostatic surface potential of Apaf-1 CARD. **c**, Stereo view of the superimposition of the structures of Apaf-1 CARD and RAIDD CARD. The structure of the RAIDD CARD domain is the best

representative conformer, as indicated in the Protein Data Bank (entry code 3CRD). r.m.s.d. values reported in the text were calculated using this NMR structure. Only 90 residues are shown for each protein, and the more divergent C-terminal residues were omitted for this comparison.

positively charged helices H1a/H1b and H4 is recognized by Apaf-1 CARD through a convex surface formed by the negatively charged helices H2 and H3 (Fig. 2), in agreement with results from our mutational analysis. Formation of the binary complex causes the burial of $\sim 1,100 \text{ \AA}^2$ of surface area.

The first two helices, H1a and H1b, appear to originate from a single, severely bent helix, and are at a crossing angle of 54 and 65 degrees, respectively, for Apaf-1 CARD and the procaspase-9 prodomain (Fig. 2). The hinge region between the H1a/H1b helices in the prodomain is intimately involved in interaction with the C-terminal end of H2 helix in the CARD.

Sequence and structure conservation

The amino-acid sequence identity between Apaf-1 CARD and the procaspase-9 prodomain is only $\sim 20\%$; however, the two structures are strikingly similar, with $1.36 \text{ \AA}$ r.m.s.d. for all C α atoms of residues 1–90 (Fig. 3). The structural similarity is evident not only in the overall topology but also in many local features. For example, three out of seven helices in Apaf-1 CARD and the procaspase-9 prodomain, H1a, H1b and H3, share the same boundaries and can be superimposed with $0.79 \text{ \AA}$ r.m.s.d. for 25 C α atoms. There are a total of six surface loops connecting the seven helices in the structure: all six loops follow the same trajectory in space for both Apaf-1 CARD and the procaspase-9 prodomain (Fig. 3b).

The sequence identity between the procaspase-9 prodomain and RAIDD CARD is 23%, but these structures exhibit significant differences, with over $3.2 \text{ \AA}$ r.m.s.d. for 90 aligned C α atoms. This difference is similar to that between the CARD domains of Apaf-1 and RAIDD.

There is little difference between the structure of free Apaf-1 CARD and that in complex with the procaspase-9 prodomain, as evidenced by the $0.41 \text{ \AA}$ r.m.s.d. for all 92 comparable C α atoms. The only notable change is that the side chains of several interface

Table 2 Mutational analysis of interaction between Apaf-1 CARD and procaspase-9 prodomain interaction			
Interaction domain	Point mutation	Apaf-1 CARD interaction	Procaspase-9 prodomain interaction
Apaf-1 CARD	Arg13Ala		++++
	Ser23Ala		++++
	Asp27Ala		----
	Ser31Ala		++/--
	Glu40Ala		----
	Lys42Ala		++++
	Lys58Ala		++++
	Lys62Ala		++++
	Leu86Ala		++++
Caspase-9 prodomain	Arg11Ala	++/--	
	Arg13Ala	----	
	Glu18Ala	++++	
	Asp23Ala	++++	
	Asp27Ala	++++	
	Leu30Ala	++++	
	Pro37Gly	++++	
	Arg52Ala	++/--	
	Arg56Ala	----	
	Asp79Ala	++++	
	Asp83Ala	++++	

Mutations that weaken CARD/prodomain interactions are in bold type, whereas mutations that abolish interactions are in bold and underlined.

residues, including those of Ser 31, Asp 27 and Glu 40, adopt different rotamer conformations in the complex to accommodate intermolecular interactions.

Electrostatic interaction at the interface

The complex formation is mediated by a negatively charged surface in Apaf-1 CARD and a highly positively charged surface in the procaspase-9 prodomain (Fig. 4). The acidic surface of Apaf-1 CARD at the interface contains three negatively charged residues, Asp 27, Glu 40 and Glu 41. The more extensive basic surface of the procaspase-9 prodomain consists of five arginine residues at positions 10, 11, 13, 52 and 56. These two surfaces are not only opposite in charge distribution but also complementary in shape, with the H1a/H1b/H4 helices of caspase-9 forming a concave surface for reception of the convex surface formed by the H2/H3 helices of Apaf-1.

Electrostatic interaction should favour the initial apposition of the two proteins, but the interaction between Apaf-1 CARD and the procaspase-9 prodomain remains undisrupted at high ionic strength (2 M NaCl; data not shown), ruling out electrostatic interaction as the driving force for mutual recognition.

Specificity of Apaf-1/procaspase-9 interaction

The specificity as well as the dominant driving force for procaspase-9 recruitment is provided by a combination of extensive hydrogen-bond contacts and van der Waals interactions (Fig. 5). There are a total of eleven intermolecular hydrogen bonds at the interface, constituting two networks.

The most extensive network involves eight hydrogen bonds and five residues, Asp 27 and Glu 40 from Apaf-1, and Arg 13, Arg 52 and Arg 56 from procaspase-9 (Fig. 5b). This network makes a large contribution to the stability of the Apaf-1/procaspase-9 complex. At the centre of this network are two anchoring residues, Asp 27 and Arg 56, which exchange a pair of charge-stabilized hydrogen bonds between their side chains (Fig. 5b). These two central contacts are capped on one side by an additional pair of charge-stabilized hydrogen bonds between the side chains of Arg 13 and Glu 40. On the other side, Arg 52 donates two hydrogen bonds to the carboxyl group of Asp 27. To strengthen this network, Arg 13 makes another intermolecular contact to the side chain of Asp 27 and Arg 56 forms an intramolecular hydrogen bond with Asp 53 (Fig. 5b).

The second network of hydrogen bonds involves Ser 31 of Apaf-1 and Arg 11 and the H1a/H1b junction of the procaspase-9 prodomain (Fig. 5c). In the structure, H1a and H1b helices are bent at a 65° crossing angle, exposing the backbone amide groups in residues 13 and 14. The side chain of Ser 31 moves into the open junction and makes two hydrogen bonds to the two amide groups, stabilizing the H1a/H1b junction and strengthening the intermolecular interaction (Fig. 5c). In addition, the carbonyl oxygen atom of

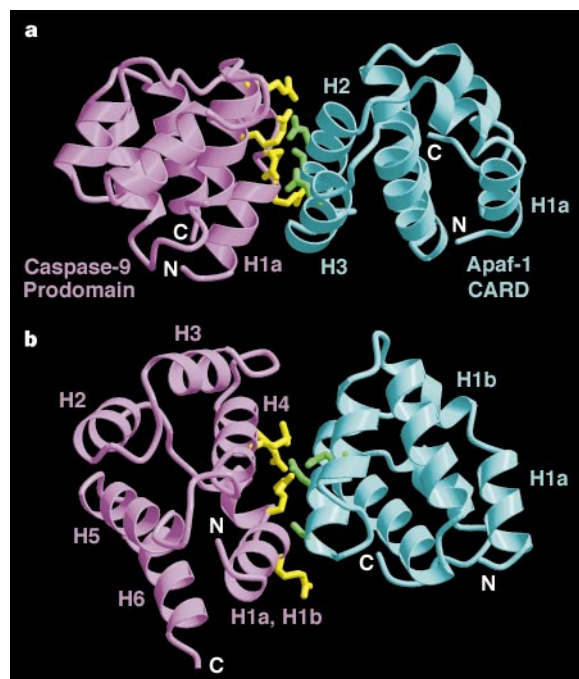


Figure 2 Representation of the structure of Apaf-1 CARD in complex with the procaspase-9 prodomain. In **a** and **b**, structures are related by a 90° rotation along the horizontal axis. Apaf-1 CARD and the procaspase-9 prodomain are coloured cyan and pink, respectively. Side chains of the important interface residues, which determine the specificity of the interaction, are in green and yellow for Apaf-1 and procaspase-9, respectively. Colour coding also applies to Figs 3 and 5; this figure and Figs 3b and 5 were prepared with MOLSCRIPT³¹.

Ser 31 makes two hydrogen bonds to the guanidinium group of Arg 11.

Electrostatic interactions and hydrogen bonds can be disrupted at high ionic strength. Because the Apaf-1 CARD/procaspase-9 prodomain interaction survives at high ionic strength, the interactions at the interface could be hydrophobic. Indeed, there is a high density of van der Waals contacts in an area adjacent to the two networks of hydrogen bonds, contributing ~30% of the 1,100 Å² total buried surface area. The core of this hydrophobic interface consists of the side chains of three residues, Ile 30 of helix H2 and Ile 37 of helix H3 in Apaf-1, and Ile 60 of helix H4 in procaspase-9 (Fig. 5d). These three residues are surrounded by the aliphatic portions of the side chains of three additional residues, Arg 10 and Arg 56 of procaspase-9, and Glu 41 of Apaf-1. To accommodate the van der Waals interactions better, the charged portion of each side chain in the

Table 3 Crystallographic analysis of Apaf-1 CARD in complex with procaspase-9 prodomain

	Reflections		Overall (outer shell)			Heavy-atom sites	MIR analysis (20–3.2 Å)		
	Resolution (Å)	Measured	Unique	Data coverage (%)	R_{sym} (%)		All (outer shell) Isomorphous difference (%)	Acentric/centric Phasing power	Cullis R factor
Native	2.5	134,080	12,452	98.6 (93.2)	5.6 (17.7)				
HgCl ₂	3.3	24,103	4,928	88.4 (92.8)	10.0 (25.5)	4	32.8 (32.3)	1.75/1.30	
Se	2.5	48,978	12,338	97.6 (89.0)	4.2 (12.6)	5	11.8 (15.1)	2.34/1.61	
K ₂ Au(CN) ₂	3.0	49,671	7,388	99.2 (100)	8.6 (25.7)	4	12.0 (17.1)	1.39/0.98	
Se/K ₂ Au(CN) ₂	3.0	49,677	7,365	99.3 (100)	9.4 (30.3)	5/3	16.2 (20.9)	2.13/1.38	
Mean figure of merit (20.0–3.2 Å): 0.689									
Refinement	Resolution	Reflections	Atoms modelled		$R_{\text{working}}/R_{\text{free}}$ (%)	r.m.s. deviation			
	range	(all)	(total, water)			Bonds (Å)	Angles (deg)	B factor (Å ²)	
	10.0–2.5	12,211	1,727, 174		22.4/29.9	0.009	1.447	4.17	

See Table 1 legend for details and calculations.

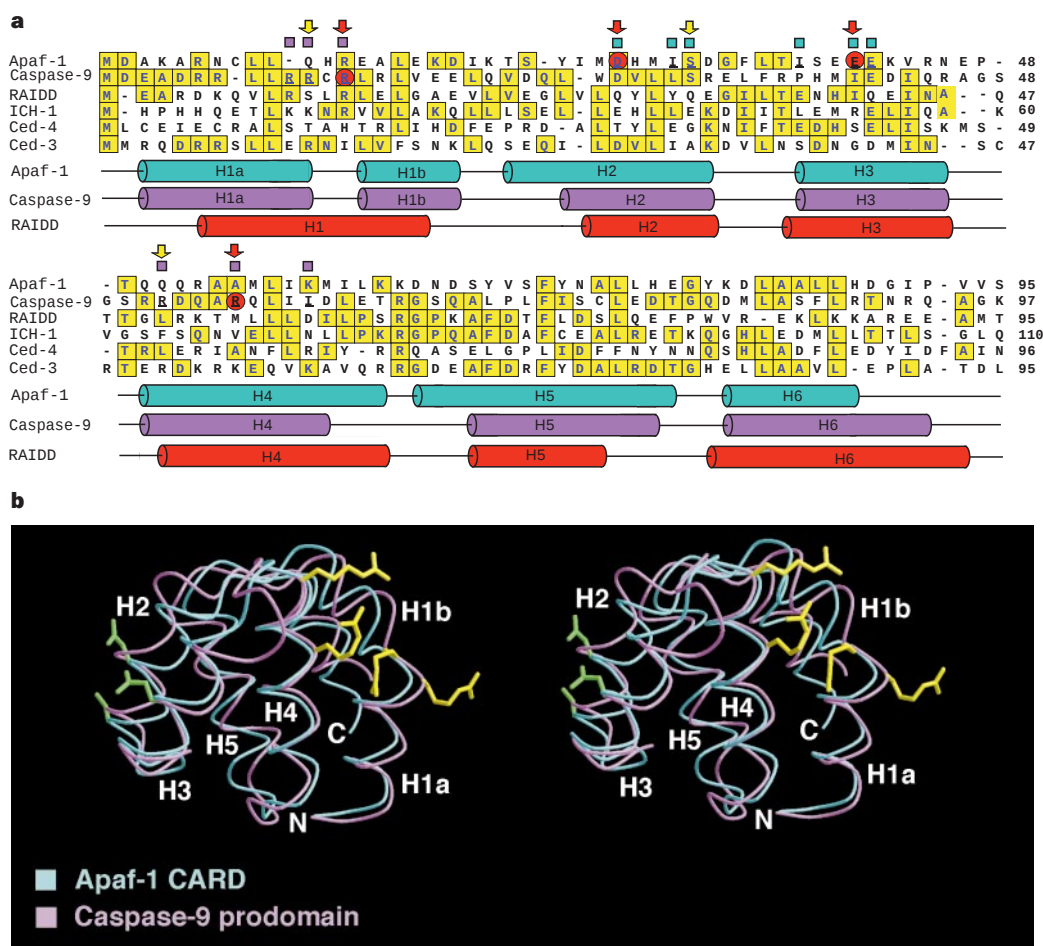


Figure 3 Sequence and structure conservation between Apaf-1 CARD and procaspase-9 prodomain. **a**, Sequence alignment of six representative CARD-homology domains. The secondary structural elements for Apaf-1 CARD, the procaspase-9 prodomain, and RAIDD CARD are indicated below the alignment. Residues involved in important interface interactions are underlined and represented by colour-coded squares above the alignment. The four residues

targeted for inactivating mutations are indicated as red circles and are identified by red arrows above the alignment; the three residues targeted for damaging mutations are indicated by yellow arrows. **b**, Superimposition of the structures of Apaf-1 CARD and procaspase-9 prodomain, shown in stereo view. To indicate the corresponding interaction surfaces in Apaf-1 and procaspase-9, the critical residues are also shown. Colour coding as in Fig. 2.

three surrounding residues points away from the hydrophobic interface (Fig. 5d).

Correlation with mutagenesis studies

The structural analysis correlates well with our mutational results (Table 2). First, all four inactivating mutations affect residues that appear to make important contributions to the stability of the interface. Asp 27 (Apaf-1), Glu 40 (Apaf-1), Arg 13 (procaspase-9) and Arg 56 (procaspase-9) lie in the centre of the interface and make 5, 2, 3 and 2 intermolecular hydrogen bonds, respectively (Fig. 5). In addition, the aliphatic portions of the side chains of Arg 13 and Arg 56 are also involved in intermolecular van der Waals contacts with adjacent hydrophobic residues. Thus, mutation of any of these four residues to alanine will result in the loss of at least two intermolecular hydrogen bonds in the centre of the interface, destabilizing the interaction and undermining the specificity for recognition.

Second, two mutations Ser31Ala of Apaf-1 and Arg52Ala of procaspase-9, weakened but did not abolish the interaction. Both wild-type residues are involved in intermolecular hydrogen bonds (Fig. 5), explaining why their mutation impairs complex stability. As both residues are located at the edge of the interface, their mutation does not prevent recognition.

Mutations and activation of procaspase-9

Catalytic activation of procaspase-9 requires assembly of the apoptosome, in which procaspase-9 is recruited by Apaf-1 in the presence of cytochrome *c* and dATP¹⁰. As alteration of the interface residues reduces or prevents binding between Apaf-1 CARD and the procaspase-9 prodomain, proteolytic processing of procaspase-9 may also be inhibited. To test this, we used a cell-free procaspase-9-activation assay, in which *in vitro* translated ³⁵S-labelled caspase-9 precursor was converted to the active form through Apaf-1-mediated autocatalytic cleavage (Fig. 6a, lane 2). Two missense mutations in procaspase-9, Arg13Ala and Arg56Ala, which disrupt recognition between Apaf-1 CARD and the procaspase-9 prodomain, blocked caspase-9 processing (Fig. 6a, lanes 3 and 7). Another mutation at the interface, Arg52Ala, which only weakened the recognition, slightly reduced the efficiency of procaspase-9 processing (Fig. 6a, lanes 6). When control mutations were introduced outside the recognition interface, mutant procaspase-9 was processed at the normal rate (Fig. 6a, lanes 4 and 5). These results agree well with our mutagenesis data and crystallographic analyses.

The amino-terminal 530 residues of Apaf-1 (Apaf-530) can induce proteolytic cleavage of caspase-9 in a cell-free system¹⁶. We did a parallel set of experiments to show that two missense

mutations in Apaf-530, Asp27Ala and Glu40Ala, fail to activate processing of procaspase-9 (Fig. 6b, lanes 4 and 6); these same mutations in the CARD domain of Apaf-1 prevent its binding to the prodomain of procaspase-9; by contrast, the Ser23Ala and Ser31Ala Apaf-530 mutants, neither of which abolishes the CARD/prodomain interaction, only decreased the amount of proteolytic cleavage of procaspase-9 (Fig. 6b, lanes 3 and 5).

Inhibition of procaspase-9 activation

Our results indicate that inhibition of caspase-9 recognition by Apaf-1 may provide a point for intervention of apoptotic pathways. We therefore investigated whether recombinant proteins could inhibit processing of procaspase-9 in the cell-free system and found that the recombinant wild-type prodomain of procaspase-9 prevented the autocatalytic processing of procaspase-9 (Fig. 6c, lane 3), whereas the inactivating mutants, Arg13Ala and Arg56Ala, had no effect (Fig. 6c, lanes 4 and 8). The damaging mutation Arg11Ala gave partial inhibition (Fig. 6c, lane 9), whereas the three mutations outside the interface (Leu30Ala, Pro37Gly and Asp27Ala) inhibited procaspase-9 activation just as the wild-type prodomain did (Fig. 6c, lanes 5, 6 and 10).

We tested the effects of recombinant Apaf-1 CARD domains on the activation of procaspase-9 (Fig. 6d) and found that the recombinant wild-type CARD domain partially inhibited proteolytic cleavage of procaspase-9 (lane 3), whereas the two inactivating mutants Asp27Ala and dGlu40Ala had no effect (lanes 5, 7). The Ser23Ala and Ser31Ala mutants, which did not prevent CARD/prodomain interaction, had little effect on the processing of procaspase-9 (Fig. 6d, lanes 4 and 6).

Our results demonstrate the feasibility of regulating apoptosis by inhibiting the activation of procaspase-9. A short isoform of caspase-9, which contains the prodomain but lacks the central catalytic domain, can act as an endogenous inhibitor for apoptosis^{26,27}; the mechanism of this inhibition was thought to involve dominant-negative interference with the formation of a functional Apaf-1–caspase-9 complex^{26,27}.

Discussion

The recognition and recruitment of CED-3 by CED-4 is likely to involve the same set of surfaces as those in the structure of Apaf-1/caspase-9, because not only are the CED-3 and CED-4 proteins of *C. elegans* the functional homologues of caspase-9 and Apaf-1, respectively, but also because the primary structures of the prodomains of CED-3 and caspase-9 are 30% identical, which is significantly higher than the identity (20%) between Apaf-1 and caspase-9. Analogous to the similarity between the structures of Apaf-1 CARD and the procaspase-9 prodomain, CED-3 is likely to retain the structural motifs and surfaces involved in recognition.

Assuming conserved backbone architecture at the interfaces, we replaced 18 residues at the Apaf-1/procaspase-9 interface with the corresponding residues from CED-4/CED-3 (Fig. 7). We then performed limited energy minimization by subjecting the mutated Apaf-1 CARD/procaspase-9 prodomain to 50 cycles of positional refinement in X-PLOR²⁸. The distribution of amino acids, both before and after energy minimization, revealed similar and interesting patterns (Fig. 7). At one end of the interface, Arg 24/Asp 25 of CED-4 interdigitates with Arg 51/Asp 52 of CED-3, possibly making several charge-stabilized hydrogen bonds; Lys 55 and Glu 56 of CED-3 may extend these contacts by additional hydrogen bonding to Asp 25, Arg 24 and Ser 45 of CED-4. At the other end of the interface, Glu 38/Asp 39 of CED-4 are within hydrogen-bonding distance of Arg 77/Glu 11 in CED-3. At the centre of the interface, Thr 28 of CED-4 is likely to have two important roles: to hydrogen-bond with Lys 55 of CED-3 and to participate in an area of dense hydrophobic interactions. These van der Waals interactions occur between the H2 helix of CED-4 and the H1a/H1b junction of CED-

3, involving residues Ile 14, Leu 15, Thr 28, Tyr 29 and the aliphatic side chain of Glu 31 (Fig. 7).

Thus, the interface between CED-4 and CED-3 is predicted to have similar features to that between Apaf-1/procaspase-9, including both networks of hydrogen bonds and patches of hydrophobic interactions. In comparison to Apaf-1/caspase-9 recognition, the identities of many important interface residues are different, but these differences correspond between CED-4 and CED-3: for example, the anchoring Asp 27 in Apaf-1 is changed to Thr 28 in CED-4, which is accompanied by a series of corresponding changes in CED-4 and CED-3. To accommodate the hydrophobic side chain of Thr 28 in CED-4, the charged residues Arg 13/Glu 14 in procaspase-9 are now replaced by Ile 14/Leu 15 in CED-3. Because of the bulky and hydrophobic nature of residues Ile 14/Leu 15 in CED-3, residue Ser 31 in Apaf-1 is now changed to Gly 32 in CED-4 to satisfy local hydrophobic packing arrangements.

Although the structural motifs at the CED-4/CED-3 interface are predicted to be conserved, the interaction specificity is determined by the divergent residues between Apaf-1/procaspase-9 and CED-4/CED-3, as evidenced by the corresponding interface residues in CED-4 and in CED-3.

With regards to structure conservation in CARDs, Apaf-1 CARD shares only 20% sequence identity with the procaspase-9 prodomain, although these two structures are nearly identical at the level of backbone atoms and can be superimposed with 1.36 Å r.m.s.d. In contrast, the structure of RAIDD CARD²⁵ varies significantly in ways that are likely to affect its function. The r.m.s.d. between RAIDD CARD and Apaf-1 CARD or the procaspase-9 prodomain is

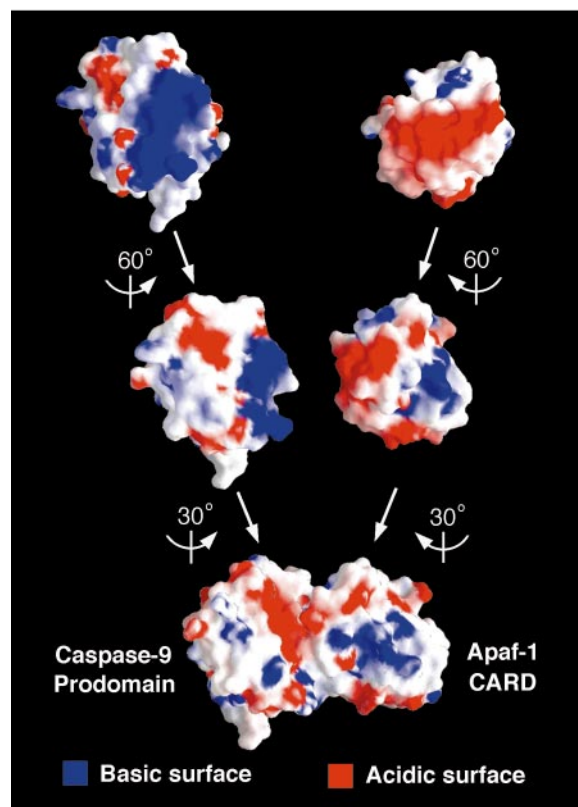


Figure 4 Electrostatic interaction at the heterodimer interface. The interaction surfaces in Apaf-1 CARD and procaspase-9 prodomain are complementary in charge distribution and in their van der Waals contour. The H1a/H1b/H4 helices of procaspase-9 present a basic and concave surface, whereas the H2/H3 helices of Apaf-1 constitute an acidic and convex surface. The electrostatic interaction between CARD and the prodomain is likely to cause the initial apposition of the two proteins. This figure was prepared with GRASP³².

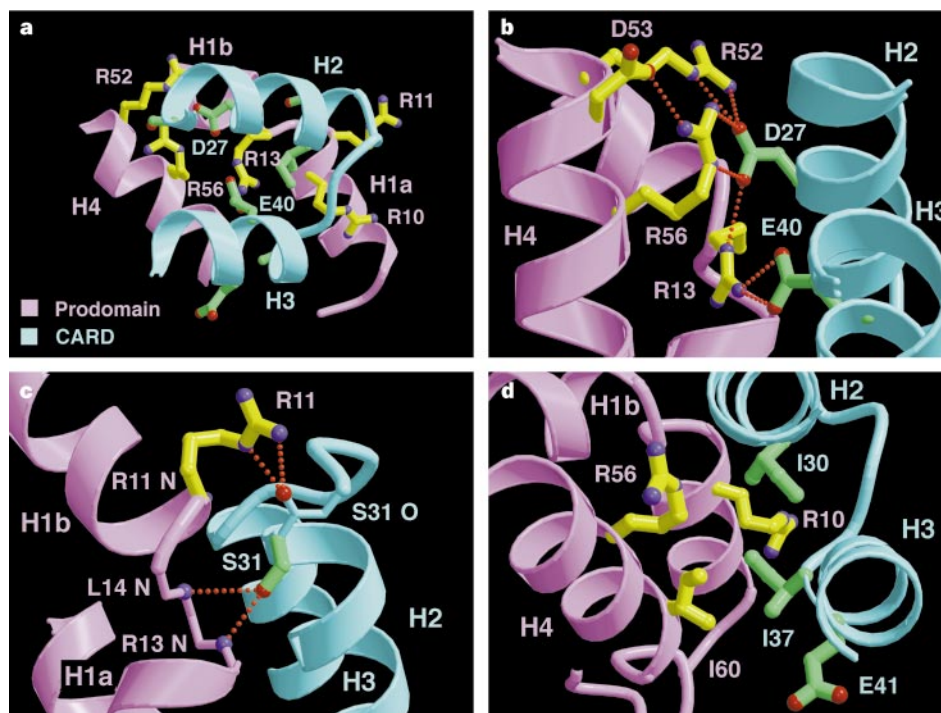


Figure 5 Recognition specificity at the heterodimer interface. **a**, In this overview, the four residues Asp27/Glu40 in Apaf-1 and Arg13/Arg56 in procaspase-9, mutation of which prevents complex formation, are at the centre of the interface. Colour coding as for Fig. 2, except that oxygen and nitrogen atoms are shown as red and blue balls, respectively. **b**, Close-up of one network of hydrogen bonds at the interface. The two anchoring residues (Asp27 and Apaf-1 and Arg56 of procaspase-9) are at the centre of this network. Hydrogen bonds are represented by red dashed lines. **c**, Close-up of a second network of hydrogen bonds involving

Ser31 of Apaf-1 and Arg11 of procaspase-9. The junction between helices H1a and H1b of procaspase-9 opens widely to accommodate the closely packed C terminus of helix H2 from Apaf-1. **d**, van der Waals interactions at the interface. The core of the hydrophobic interface is formed by Ile30 and Ile37 from Apaf-1 and by Ile60 from procaspase-9, with additional contacts from the aliphatic portions of the side chains of three charged residues—Glu41 of Apaf-1, and Arg10 and Arg56 of procaspase-9.

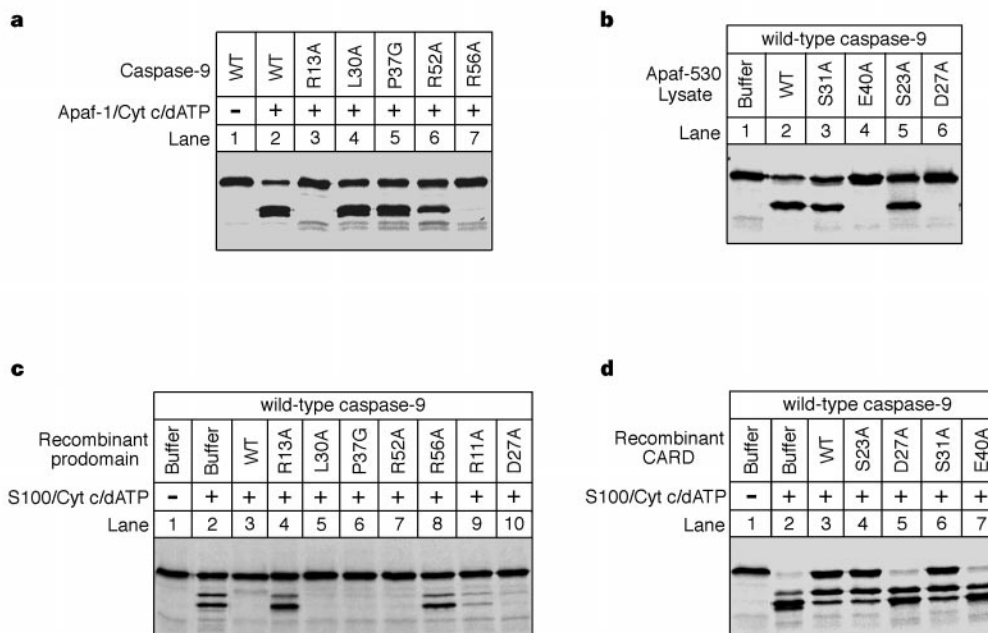


Figure 6 Effect of point mutations on the catalytic activation of procaspase-9. **a**, Effect of point mutations in procaspase-9. Procaspase-9 wild-type (WT) and mutant precursor molecules were translated *in vitro* in the presence of ^{35}S -methionine. Recombinant Apaf-1 was expressed from baculovirus-infected insect cells and purified through a Ni^{2+} -affinity resin. **b**, Effect of point mutations in Apaf-530. Although Apaf-530 can induce proteolytic processing of procaspase-9 in the absence of cytochrome c and dATP (ref. 16), both cytochrome c (5 ng μm^{-1})

and dATP (1 mM) were provided for all reactions. **c**, Inhibition of catalytic activation of procaspase-9 by recombinant procaspase-9 prodomain. Equal amounts (2 μg) of wild-type or mutant prodomain of procaspase-9 (residues 1–132) were used for each lane. **d**, Inhibition of catalytic activation of procaspase-9 by recombinant Apaf-1 CARD domain. Equal amounts (2 μg) of wild-type or mutant Apaf-1 CARD domain (residues 1–97) were used for each reaction.

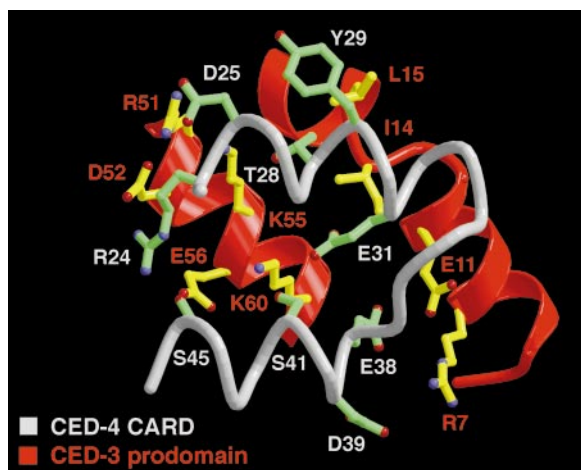


Figure 7 Predicted recognition interface between CED-3 and CED-4. The overall architecture of CED-4/CED-3 interface is shown in the same orientation as that of Apaf-1/procaspase-9 in Fig. 5a. The H2 and H3 helices of CED-4 are represented as grey coils for better visualization of the interface residues. The H1a/H1b/H4 helices of CED-3 are in red; residues at the interface are in green and yellow for CED-4 and CED-3, respectively.

over 3.2 Å. The large differences can be explained only in part by the fact that eight surface loop residues were not assigned in the RAIDD CARD structure²⁵. After removal of all loop residues from the calculation, the r.m.s.d. for the remaining 74 Cα atoms (2.5 Å) is still significantly higher than that between Apaf-1 CARD and the procaspase-9 prodomain (1.1 Å). The structural differences cannot be explained by the extent of sequence identity or similarity, because in both categories caspase-9 resembles RAIDD more closely than Apaf-1. Finally, the differences cannot be attributed to crystal packing or to conformational changes associated with complex formation, because the structure of free Apaf-1 CARD can be superimposed on that in the complex with 0.41 Å r.m.s.d.

One explanation for the structural differences is functional variation. RAIDD may act in the TNF-R1-induced signalling pathway by recruiting caspase-2 (ICH-1) through CARD/prodomain recognition²⁵. We have been unable to model the molecular interactions between RAIDD CARD and the caspase-2 prodomain on the basis of the structure reported here.

The recognition of caspase-9 by Apaf-1 is highly specific, relying on two networks of hydrogen bonds and a surface patch enriched with intermolecular van der Waals interactions. Mutations of key residues at the interface cause complex formation to fail (Table 2). The specificity of CARD/prodomain recognition is further demonstrated by the Apaf-1/caspase-9 functional homologues in *C. elegans*, the CED4/CED-3 complex. None of the four residues targeted for mutational inactivation in the Apaf-1/procaspase-9 interaction is conserved in CED-4/CED-3, yet the corresponding changes in the predicted CED-4/CED-3 interface allow formation of a novel set of intermolecular hydrogen bonds and van der Waals interactions.

Because only one charged surface in each domain is used for complex formation, the remaining basic surface of Apaf-1 CARD and the acidic surface of the procaspase-9 prodomain could contribute to additional interdomain interactions such as oligomerization. However, under all experimental conditions examined, the only detectable complex was a heterodimer, as observed in the structure. In addition, mutations targeted at the remaining two surfaces had no effect on the formation of a binary complex (Table 2), emphasizing the interaction specificity. The oligomerization of the Apaf-1/procaspase-9 complex is mediated by the CED-4-homology domain present in Apaf-1 (refs 16, 17).

Recruitment of procaspase-9 by Apaf-1 must be central to its

subsequent activation as disruption of this process causes complete inhibition of procaspase-9 processing in a cell-free system (Fig. 6). In addition, recombinant prodomain proteins can block the activation of procaspase-9 (Fig. 6c), raising the possibility of using small molecules to intervene in apoptotic pathways, an idea that is supported by the rich surface features of Apaf-1 and procaspase-9, which may help in the identification of small molecules with high binding affinity. □

Methods

Site-directed mutagenesis and protein preparation. Point mutations were generated using a standard PCR-based cloning strategy, and the identities of individual clones were verified by double-strand plasmid sequencing. Recombinant caspase-9 prodomain (residues 1–112) was overexpressed in *Escherichia coli* strain BL21(DE3) as a glutathione-S-transferase (GST) fusion protein using a pGEX-2T vector (Pharmacia). The soluble fraction of the GST–procaspase-9 prodomain fusion in the *E. coli* lysate was purified over a glutathione–Sepharose column, cleaved by thrombin, and further purified by anion-exchange chromatography (Source 15Q; Pharmacia) and gel-filtration chromatography (Superdex-75; Pharmacia). Recombinant Apaf-1 CARD (residues 1–97) was likewise overexpressed using a pET-3d vector (Novagen). Selenomethionyl CARD was expressed in *E. coli* B834(DE3) (Novagen) in M9 minimal medium supplemented with 50 mg l⁻¹ selenomethionine. The soluble fraction of the Apaf-1 CARD was purified by anion-exchange and gel-filtration chromatography.

The concentrations of Apaf-1 CARD and caspase-9 prodomain were determined by spectroscopic measurement at 280 nm. Equimolar amounts of these two proteins were combined, concentrated, and further purified through a Superdex-75 column (Pharmacia). The final concentration of the CARD/prodomain complex was ~20 mg ml⁻¹.

In vitro interaction assay. Six-exclusion chromatography on a Superdex-75 column (Pharmacia) was used to examine the Apaf-1 CARD/caspase-9 prodomain interaction. The flow rate was 0.5 ml min⁻¹ and the buffer contained 25 mM Tris, pH 8.0, 150 mM NaCl, and 2 mM DTT.

Crystallization and data collection for Apaf-1 CARD. Crystals were grown at 22 °C by the hanging-drop vapour-diffusion method by mixing Apaf-1 CARD protein (20 mg ml⁻¹) with an equal volume of reservoir solution containing 100 mM acetate buffer, pH 4.6, 24–28% PEG4000 (w/w), 200 mM ammonium acetate, and 10 mM DTT. The crystals, with a typical size of 0.25 × 0.25 × 0.80 mm³, were in the monoclinic space group C2, with unit-cell dimensions of *a* = 56.7 Å, *b* = 36.0 Å, *c* = 47.9 Å, β = 118.2 degrees, and contain one protein in the asymmetric unit. Diffraction data were collected using an R-AXISIIIC imaging plate detector mounted on a Rigaku 200HB generator. Derivatives were obtained by soaking crystals in the appropriate heavy-atom solutions in 100 mM acetate buffer, pH 4.6, and 25% PEG4000 (w/w). Concentrations and soaking times for HgCl₂, K₂PtCl₄ and K₂AuCl₂ were 0.5 mM for 10 h, 0.5 mM for 11 h, and 1 mM for 12 h, respectively. For the frozen native data set, crystals were equilibrated in a cryoprotectant buffer containing 100 mM acetate, pH 4.6, 25% PEG4000 (w/w), 20% glycerol, and were flash-frozen under a -170 °C nitrogen beam.

Structure determination for Apaf-1 CARD. The heavy-atom position of the HgCl₂ derivative was determined by the difference Patterson method; all other derivatives were identified by the difference Fourier method. Initial MIR phases calculated with the program MLPHARE²⁹ had a mean figure of merit of 0.855 to 2.1 Å, and were improved with solvent flattening and histogram matching with the program DM (ref. 29). A model was built into MIR electron density maps with the program O (ref. 30) and refined by simulated annealing with the program X-PLOR²⁸. The final refined model contains Apaf-1 residues 1–92 and 138 water molecules. The C-terminal five residues in Apaf-1 CARD have no electron density in the maps, and we presume that these regions are disordered in the crystals.

Mutagenesis of Apaf-530 and full-length procaspase-9. Both wild-type and mutant Apaf-530 were cloned into bacterial expression vectors pET-28a (Invitrogen); genes encoding wild-type and mutant procaspase-9 in the mammalian expression vector pRSC-LacZ were used for *in vitro* translation¹⁶.

Crystallization and data collection for the complex. Crystals were grown at 4 °C by the hanging-drop vapour-diffusion method by mixing the complex

with an equal volume of reservoir solution containing 100 mM HEPES buffer, pH 7.5, 1.75 M potassium sodium dihydrogen phosphate, 10 mM DTT. The crystals, with a typical size of $0.25 \times 0.25 \times 0.12 \text{ mm}^3$, were in the hexagonal space group $P6_22_2$, with unit-cell dimensions of $a = b = 92.5 \text{ \AA}$, $c = 136.8 \text{ \AA}$, $\gamma = 120$ degrees, and contain one complex in the asymmetric unit. Crystals were equilibrated in a cryoprotectant buffer containing 100 mM HEPES, pH 7.5, 2.0 M potassium sodium dihydrogen phosphate, and 20% glycerol, and were flash-frozen under a -170°C nitrogen beam. Both native and derivative data sets were collected at -170°C . Derivatives were obtained by soaking crystals with appropriate heavy atoms in 100 mM HEPES buffer, pH 7.5, 2.25 M potassium sodium dihydrogen phosphate. The concentration and soaking time for HgCl_2 and $\text{K}_2\text{Au}(\text{CN})_2$ were 0.2 mM for 8 h and 1 mM for 34 h, respectively.

Structure determination for the complex. The first heavy-atom position of the HgCl_2 derivative was determined by the difference Patterson method, and the second heavy-atom site and those of all other derivatives were identified by the difference Fourier method. Initial MIR phases calculated with the program MLPHARE²⁹ had a mean figure of merit of 0.687 to 3.2 Å and were improved by solvent flattening and histogram matching with the program DM²⁹. A model was built into MIR electron density maps with program O³⁰ and refined with program X-PLOR²⁸. The final refined model contains Apaf-1 residues 1–95, procaspase-9 residues 1–97, and 174 water molecules. The C-terminal two residues in Apaf-1 CARD and residues 98–112 in procaspase-9 had no electron density in the maps; we presume that these regions are disordered in the crystals.

Reconstitution of procaspase-9 activation. Experiments were done as described¹⁶.

Received 21 April; accepted 7 May 1999.

- Steller, H. Mechanisms and genes of cellular suicide. *Science* **267**, 1445–1449 (1995).
- Jacobson, M. D., Weil, M. & Raff, M. C. Programmed cell death in animal development. *Cell* **88**, 347–354 (1997).
- Hengartner, M. O. Programmed cell death in invertebrates. *Curr. Opin. Genet. Dev.* **6**, 34–38 (1996).
- Horvitz, H. R. Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res.* **59**, 1701–1706 (1999).
- Thompson, C. B. Apoptosis in the pathogenesis and treatment of disease. *Science* **267**, 1456–1462 (1995).
- Green, D. R. & Martin, S. J. The killer and the executioner: how apoptosis controls malignancy. *Curr. Opin. Immunol.* **7**, 694–703 (1995).
- Thornberry, N. A. & Lazebnik, Y. Caspases: enemies within. *Science* **281**, 1312–1316 (1998).
- Chinnaiyin, A. M. & Dixit, V. M. The cell-death machine. *Curr. Biol.* **6**, 555–562 (1996).
- Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to *C. elegans* CED-4, participates in cytochrome c-dependent activation of caspase-3. *Cell* **90**, 405–413 (1997).
- Li, P. *et al.* Cytochrome c and dATP-dependent formation on Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479–489 (1997).
- Yoshida, H. *et al.* Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**, 739–750 (1998).
- Cecconi, F., Alvarez-Bolado, G., Meyer, B. L., Roth, K. A. & Gruss, P. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
- Soengas, M. S. *et al.* Apaf-1 and Caspase-9 in p53-dependent apoptosis and tumor inhibition. *Science* **284**, 156–159 (1999).
- Fearnhead, H. O. *et al.* Oncogene-dependent apoptosis is mediated by caspase-9. *Proc. Natl. Acad. Sci.* **95**, 13664–13669 (1998).
- Slee, E. A. *et al.* Ordering the cytochrome c-initiated caspase cascade: hierarchical activation of caspase-2, -3, -6, -7, -8, and -10 in a caspase-9-dependent manner. *J. Cell. Biol.* **144**, 281–292 (1999).
- Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T. & Alnemri, E. S. Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. *Mol. Cell* **1**, 949–957 (1998).
- Yang, X., Chang, H. Y. & Baltimore, D. Essential role of CED-4 oligomerization in CED-3 activation and apoptosis. *Science* **281**, 1355–1357 (1998).
- Hu, Y., Ding, L., Spencer, D. M. & Nunez, G. WD-40 repeat region regulates Apaf-1 self-association and procaspase-9 activation. *J. Biol. Chem.* **273**, 33489–33494 (1998).
- Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1-cytochrome c multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).
- Saleh, A., Srinivasula, S. M., Acharya, S., Fishel, R. & Alnemri, E. S. Cytochrome c and dATP-mediated oligomerization of Apaf-1 is a prerequisite for procaspase-9 activation. *J. Biol. Chem.* (in the press).
- Hofmann, K. & Bucher, P. The CARD domain: a new apoptotic signaling motif. *Trends Biochem. Sci.* **22**, 155–156 (1997).
- Duan, H. & Dixit, V. M. RAIDD is a new 'death' adaptor molecule. *Nature* **385**, 86–89 (1997).
- Ahmad, M. *et al.* CRADD, a novel human apoptotic adaptor molecule for caspase-2, and FasL/tumor necrosis factor receptor-interacting protein RIP. *Cancer Res.* **57**, 615–619 (1997).
- Alnemri, E. S. *et al.* Human ICE/CED-3 protease nomenclature. *Cell* **87**, 171 (1996).
- Chou, J. J., Matsuo, H., Duan, H. & Wagner, G. Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. *Cell* **94**, 171–180 (1998).
- Seol, D.-W. & Billiar, T. R. A caspase-9 variant missing the catalytic site is an endogenous inhibitor of apoptosis. *J. Biol. Chem.* **274**, 2072–2076 (1999).
- Srinivasula, S. M. *et al.* Identification of an endogenous dominant-negative short isoform of caspase-9 that can regulate apoptosis. *Cancer Res.* **59**, 999–1002 (1999).
- Brunger, A. T. *X-PLOR, a System for Crystallography and NMR* (Yale University Press, New Haven, CT, 1991).
- Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Klaulis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Acknowledgements. We thank S. Kyin for peptide sequencing and mass spectroscopic analysis; F. Hughson for critically reading the manuscript; A. Saleh for purifying recombinant Apaf-1; R. Mukattash for technical assistance; and E. Steckman for secretarial assistance. This work was supported by start-up funds from Princeton University (to Y.S.) and grants from the NIH (to E.S.A.). Y.S. is a Searle Scholar and a Rita Allen Scholar.

Correspondence and requests for materials should be addressed to Y.S. (e-mail: ygshi@princeton.edu). Coordinates have been deposited with the Protein Data Bank (accession numbers 2ygs and 3ygs).

Detection of an impact-generated dust cloud around Ganymede

Harald Krüger*, Alexander V. Krivov†, Douglas P. Hamilton‡ & Eberhard Grün*

* Max-Planck-Institut für Kernphysik, Postfach 103980, 69029 Heidelberg, Germany

† Astronomical Institute, St Petersburg University, 198904 St Petersburg, Russia

‡ University of Maryland, College Park, Maryland 20742-2421, USA

Dust pervades the Solar System, and is concentrated in the ring systems surrounding the giant planets and along the plane of the planetary orbits (the Zodiacal cloud). Individual dust grains are thought to be generated when impacts loft material from larger bodies^{20,21,23–27}, such as satellites. Uncertainties in theoretical models of this ejection process are large, and there have hitherto been no direct measurements with which to constrain these models. Here we report *in situ* measurements of submicrometre dust within a few radii of Jupiter's satellite Ganymede. The directions, speeds and distribution of masses of the grains indicate that they come from Ganymede, and are consistent with an ejection process resulting from hypervelocity impacts of interplanetary dust onto Ganymede's surface. Dust appears also to be concentrated near Callisto and Europa, suggesting that these satellites too are significant sources of dusty debris.

The Galileo spacecraft, orbiting Jupiter since 7 December 1995, is equipped with an impact ionization dust detector^{1,2}, which measures the plasma cloud released on impact of submicrometre and micrometre dust particles onto its sensor. Masses and impact speeds of the grains are determined from the measured amplitudes and rise times of the impact charge signals in three channels³. Each impact event is classified into one of four quality classes, with class 3 being dust impacts and class 0 being noise. Depending on the noise rate of the charge measurements on the individual channels, classes 1 and 2 can be true dust impacts or noise events^{3,4}. As class 2 events were relatively noise-free during the Ganymede fly-bys, here we analyse the combined class 2 and class 3 data set.

During close fly-bys of the galilean moons, the overall impact rate of dust grains measured by the dust sensor showed a sharp peak within about half an hour centred on closest approach to each satellite^{5–7}. Concentrations of dust were observed for Europa, Ganymede and Callisto, with the Ganymede data set being the most complete; during the combined primary and Galileo Europa Mission (GEM), Galileo had four close fly-bys at Ganymede, and

no further encounters with this satellite are planned in the future. The times of the Ganymede fly-bys are given in Table 1, and the geometry of the Galileo dust detections is explained in Fig. 1.

The direction in which the sensor was pointing during dust impacts (rotation angle, Θ) at the Ganymede encounters is shown in Fig. 2. During the first three encounters (G1, G2 and G7) particles with $180^\circ \leq \Theta < 360^\circ$ were detected at altitudes below $4R_G$ (Ganymede radius, $R_G = 2,635$ km) and were concentrated towards Ganymede. Particles recorded from the opposite direction ($0^\circ \leq \Theta < 180^\circ$) did not show such a concentration⁶.

Analysis of the velocity vector of Ganymede relative to Galileo, taking into account the 140° field of view of the dust detector¹, shows that particles belonging to a dust cloud around Ganymede could be detected with $\Theta = 270 \pm 90^\circ$ during all four encounters. Thus particles with $180^\circ \leq \Theta \leq 360^\circ$ detected during the first three fly-bys are compatible with having a Ganymede origin. We will call them Ganymede particles. Particles detected from the opposite direction ($0^\circ \leq \Theta < 180^\circ$) are streams of 10-nm dust grains^{8,9} which probably originate from Io (refs 6, 7, and A. L. Graps *et al.*, manuscript in preparation). They are not considered here.

We have identified 35 Ganymede particles from the first three encounters purely by their impact direction (Table 1). The measured impact speeds of most Ganymede particles from these encounters

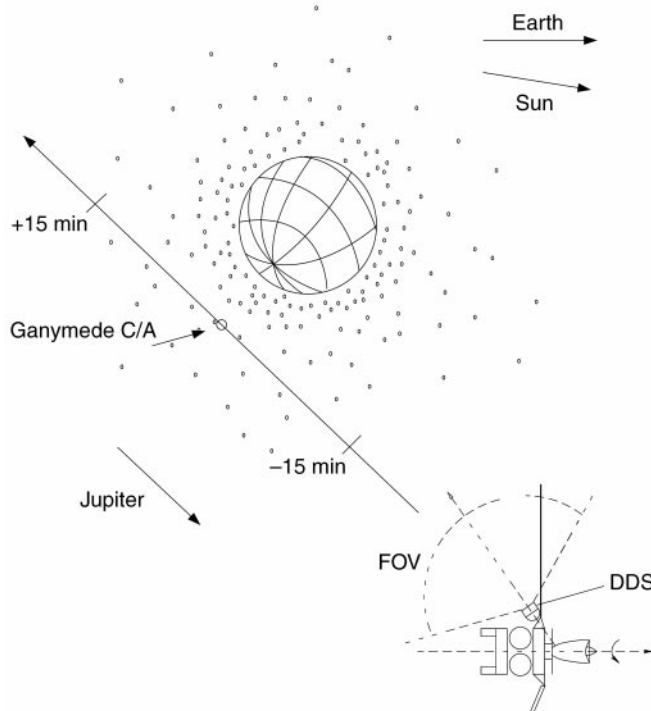


Figure 1 Galileo's trajectory and geometry of dust detection during the G7 Ganymede fly-by. The Galileo spacecraft is sketched in an orientation it was in during the fly-by: the antenna points towards Earth and the dust detector scans the opposite hemisphere away from the Earth. The dust detector (DDS) has a conical field of view (FOV) of 140° and is mounted at an angle of 60° with respect to the positive spin-axis (anti-Earth direction). As Galileo spins about the spacecraft–Earth line, the dust detector axis sweeps out a cone with 120° opening angle, sampling dust arriving from different directions. The dust detector is shown here in an orientation where particles belonging to a cloud of dust from Ganymede can be detected (rotation angle $\Theta \approx 270^\circ$). Dust-stream particles approach from the opposite direction (Jupiter direction, $\Theta \approx 90^\circ$). C/A, closest approach.

Table 1 Satellite encounter characteristics and dust impact detections

Fly-by	Time of C/A (year-day)	Altitude at C/A (km)	Particles with full data set	Corrected number of particles
G1	96-179.270	844	15	30
G2	96-250.791	262	9	48
G7	97-095.299	3,095	11	11
G8	97-127.665	1,596	9*	49

Galileo satellite fly-bys (first column) are labelled with the first letter of the galilean satellite which was the encounter target during that orbit, followed by the orbit number. C/A, closest approach. Galileo's data transmission capability is very limited because its high-gain antenna failed to open completely. Although each impact event is counted, the full set of measured parameters (such as impact direction, impact charge, and charge rise times) is not always transmitted to Earth. During the satellite encounters, the transmission rate was roughly one full impact parameter set per minute. The higher impact rates experienced during the G1, G2 and G8 encounters resulted in loss of some data. Because all impacts were counted, however, we can derive the true number of Ganymede particles (column 5) by multiplying the total number of impacts (from all directions) by the fraction of Ganymede particles for which we have full data sets (column 4)¹⁰.

* Only particles with impact speed $v \leq 10$ km s⁻¹ and below $5R_G$ altitude included.

are below 10 km s^{-1} with a mean value of $7.0 \pm 4.8 \text{ km s}^{-1}$ (ref. 10). The stream particles, by contrast, have typical impact speeds (derived from the instrument's calibration) which are much higher than 10 km s^{-1} . Although the instrument calibration probably significantly underestimates the true speeds of the tiny dust-stream particles¹¹, the calibrated impact speeds can be used to distinguish Ganymede particles from the dust streams (Fig. 2, bottom panel). The measured impact speeds of Ganymede particles agree well with the expected impact speeds for debris moving with Ganymede, $\sim 8 \text{ km s}^{-1}$ for all encounters, which suggests that these grains do indeed originate from Ganymede.

During Galileo's fourth Ganymede fly-by (G8), both dust streams and Ganymede particles approached the sensor from the same direction ($180^\circ \leq \theta < 360^\circ$). On the basis of data from the first three encounters, we adopt two criteria to distinguish Ganymede particles from dust streams: (1) Ganymede particles must be within $5R_G$ of the satellite, and (2) the measured impact speed must be $< 10 \text{ km s}^{-1}$. Nine particles satisfy these criteria (Table 1).

To characterize the dust cloud of Ganymede, we derived both mass and spatial distributions for the detected grains. The slope of the cumulative mass distribution for particles in the mass range 10^{-16} to 10^{-13} kg is $\alpha = 0.98 \pm 0.06$ (ref. 10). This is consistent with the typical slopes expected for ejecta ($0.5 \leq \alpha \leq 1.0$; refs 12, 13).

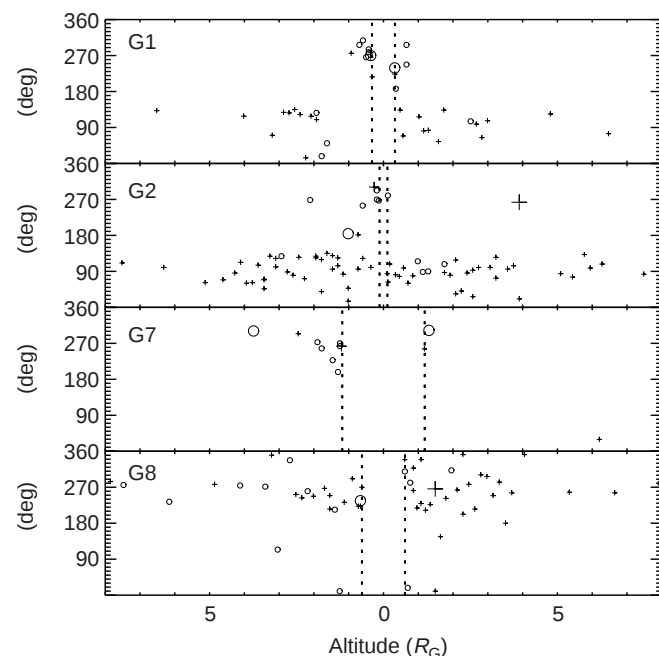


Figure 2 Sensor direction (rotation angle, θ) versus altitude of the Galileo spacecraft above the surface of Ganymede at the time of dust impact. Data are shown for all four Ganymede encounters (G1, G2, G7, G8). The altitude range shown corresponds to a time interval of 1.6 h. At $\theta = 0^\circ$ the sensor axis points closest to ecliptic north, at 90° it points closest to the direction of Jupiter. The direction to Ganymede is $\sim 270^\circ$ during approach. Here we plot only impacts for which we have a complete set of parameters. The apparent concentration of these particles within $5R_G$ is due to an increased data transmission rate near Ganymede. Circles show particles with impact speeds below 10 km s^{-1} and crosses show particles with higher speeds. The symbol sizes indicate the impact charge created by the particles ($10^{-14} \text{ C} \leq Q_i \leq 10^{-11} \text{ C}$). Ganymede's radius is $R_G = 2,635 \text{ km}$. Galileo did not pass through the altitude ranges between the dotted lines. For G1, G2 and G7, Ganymede particles approached from the opposite direction ($180^\circ < \theta \leq 360^\circ$) to the stream particles ($0^\circ < \theta \leq 180^\circ$). In G8, both types of particles approached from the same direction and they can be distinguished by their impact speed only.

The spatial distribution of the dust grains is shown in Fig. 3. During the first three encounters (G1, G2 and G7) the number density increases towards Ganymede, with power-law slopes ranging between -0.43 and -2.85 . For the G8 encounter, no concentration of particles towards Ganymede is seen, perhaps due to uncertainties imposed by the separation from the stream particles and the large correction for incomplete transmission. The concentration of dust towards Ganymede leaves no doubt that the satellite itself is the source, as its gravitational and electromagnetic forces are too weak to appreciably focus interplanetary and/or interstellar dust. Because there are no indications of geysers or volcanoes on Ganymede, the most likely source is the continuous ejection of debris via bombardment of Ganymede's surface by interplanetary micrometeoroids.

In Fig. 3, we combine the data from all four Ganymede fly-bys on a single plot. In interpreting the plot, we assume that the flux of dust released from Ganymede is constant in time and is constant over the satellite's surface. These assumptions are probably not entirely true, but are accurate enough to allow simple order-of-magnitude estimates to be made. Most ejecta grains follow ballistic trajectories and strike Ganymede again within several hours to a few days. These short-lived but continuously replenished grains form a tenuous steady-state debris cloud which entirely envelopes Ganymede. Our measurements suggest a surprisingly large amount of orbiting ejecta

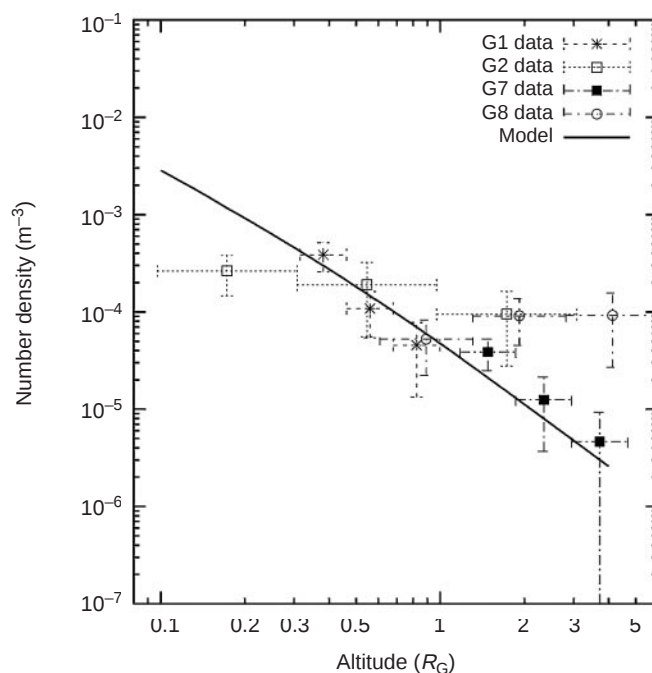


Figure 3 The number density of dust as a function of altitude above the surface of Ganymede. To obtain the number density from the data (symbols with error bars), we first defined altitude ranges ('bins') equally spaced outwards from Ganymede on a logarithmic scale (indicated by horizontal bars). We then divided the number of particles for which the complete set of parameters has been transmitted to Earth in a given distance bin by the time Galileo has spent in this bin. We corrected for incomplete data transmission (Table 1) and divided the rates by the effective spin-averaged detector area to obtain fluxes (in $\text{m}^{-2} \text{ s}^{-1}$). Finally, we divided the results by the mean impact velocity for a given fly-by, which results in mean number densities (in m^{-3}) in various bins. Vertical error bars reflect statistical errors due to the small number of impacts. The solid curve is the theoretical distribution of the impact ejecta expected for interplanetary impactors with a plausible set of model parameters^{10,14,15}.

in Ganymede's debris cloud: a steady-state value of roughly ten tons with a factor of 10 uncertainty.

The solid curve in Fig. 3 shows the predictions of an impact ejecta model^{10,14,15}, in which the interplanetary dust particles strike the surface of Ganymede, ejecting secondary particles. After choosing plausible values for the flux of interplanetary micrometeoroids and the physical properties of Ganymede's surface, we find that the model predictions fit the dust data reasonably well. Our measurements and dynamical modelling therefore strongly suggest that the dusty debris near Ganymede is produced by a continuous hail of interplanetary particles which strike this moon with enough energy to accelerate dusty debris off its surface. Similar clouds probably surround Europa and Callisto, and, indeed, any moon that lacks a gaseous atmosphere. The few previous attempts to directly detect dust close to satellites, most notably near the Moon¹⁶, have led to inconclusive results. Our successful detection of dust in the vicinity of the large jovian satellites underscores the general nature of the process, and provides strong support that our own Moon is a source of dust in near-terrestrial space.

A tiny fraction of impact debris is ejected at speeds sufficient to escape from Ganymede entirely. This material goes into orbit around Jupiter and will eventually be swept up by one of the galilean satellites—these grains are probably responsible for some of the impact events detected by the dust instrument in the inner jovian system^{5,6}. Unfortunately, the ring of material formed by these grains escaping from Ganymede is far too tenuous to be detected optically. However, the fraction of debris escaping a satellite is a steeply decreasing function of satellite mass: so steep that despite their reduced cross-sections, small moons may be better sources of dust than large moons¹⁷. This is most clearly exemplified in the images of the jovian ring provided by Galileo. These images show that the two tiny innermost moons, Adrastea and Metis, are sources for the main jovian ring and dusty halo, while Thebe and Amalthea each give rise to a faint outer gossamer ring of dusty material^{18,19}. Medium-sized moons can also produce detectable rings: at Saturn the moon Enceladus supplies material to the broad and tenuous E ring²⁰ which, nevertheless, is substantial enough to be visible to ground-based telescopes. Impact ejection of dusty debris is likely to be important for explaining the distinctive black–white asymmetry of Saturn's moon Iapetus²¹, for transporting exogenous material to Titan^{21,22}, and for producing dust in the uranian²³ and neptunian ring systems²⁴.

Future spacecraft measurements near planetary satellites are expected to quantify this important dust-production mechanism, and provide critical insight into the impact genesis of dusty rings. The next spacecraft likely to make *in situ* measurements of impact-generated grains is Nozomi (formerly called Planet-B). It will attempt to detect dust belts around Mars, which are predicted to originate from impacts onto Phobos and Deimos^{25,26}. In mid-2004, the Cassini spacecraft should arrive at Saturn. Cassini, which is equipped with an improved version of the Galileo dust instrument, is intended to make numerous passes near the saturnian satellites and will fly through the tenuous E ring, thereby sampling both the bound and escaping components of Enceladus ejecta. □

Received 12 February; accepted 15 April 1999.

1. Grün, E. *et al.* The Galileo dust detector. *Space Sci. Rev.* **60**, 317–340 (1992).
2. Grün, E. *et al.* The Ulysses dust experiment. *Astron. Astrophys. Suppl. Ser.* **92**, 411–423 (1992).
3. Grün, E. *et al.* Reduction of Galileo and Ulysses dust data. *Planet. Space Sci.* **43**, 941–951 (1995).
4. Krüger, H. *et al.* Three years of Galileo dust data: II. 1993 to 1995. *Planet. Space Sci.* **47**, 85–106 (1999).
5. Grün, E. *et al.* Dust measurements in the jovian magnetosphere. *Geophys. Res. Lett.* **24**, 2171–2174 (1997).
6. Grün, E. *et al.* Galileo observes electromagnetically coupled dust in the jovian magnetosphere. *J. Geophys. Res.* **103**, 20011–20022 (1998).
7. Krüger, H., Grün, E., Graps, A. & Lammers, S. Observations of electromagnetically coupled dust in the Jovian magnetosphere. *Astrophys. Space Sci.* (in the press).
8. Grün, E. *et al.* Discovery of jovian dust streams and interstellar grains by the Ulysses spacecraft. *Nature* **362**, 428–430 (1993).
9. Grün, E. *et al.* Constraints from Galileo observations on the origin of jovian dust streams. *Nature* **381**, 395–398 (1996).
10. Krüger, H., Krivov, A. V., Grün, E. & Hamilton, D. P. A dust cloud of Ganymede maintained by hypervelocity impacts of micrometeoroids. *J. Geophys. Res.* (submitted).

11. Zook, H. A. *et al.* Solar wind magnetic field bending of jovian dust trajectories. *Science* **274**, 1501–1503 (1996).
12. Koschny, D. & Grün, E. Impacts into ice-silicate mixtures: Crater morphologies, volumes, depth-to-diameter ratios and yield. *Icarus* (submitted).
13. Koschny, D. & Grün, E. Impacts into ice-silicate mixtures: Ejecta mass and size distributions. *Icarus* (submitted).
14. Krivov, A. V. On the dust belts of Mars. *Astron. Astrophys.* **291**, 657–663 (1994).
15. Krivov, A. V. & Jurewicz, A. The ethereal dust envelopes of the Martian moons. *Planet. Space Sci.* **47**, 45–56 (1999).
16. Iglöeder, H., Uesugi, K. & Svedhem, H. Cosmic dust measurements in lunar orbit. *Adv. Space Res.* **17**, 177–182 (1996).
17. Burns, J. A., Showalter, M. R. & Morfill, G. E. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 200–272 (Univ. Arizona Press, Tucson, 1984).
18. Ockert-Bell, M. E. *et al.* The structure of Jupiter's ring system as revealed by the Galileo imaging experiment. *Icarus* **138**, 188–219 (1999).
19. Burns, J. A. *et al.* The formation of Jupiter's faint rings. *Science* **284**, 1146–1150 (1999).
20. Hamilton, D. P. & Burns, J. A. Origin of Saturn's E ring: self-sustained, naturally. *Science* **264**, 550–553 (1994).
21. Burns, J., Hamilton, D. P., Mignard, F. & Soter, S. in *Physics, Chemistry, and Dynamics of Interplanetary Dust* (eds Gustafson, B. A. S. & Hanner, M. S.) 179–182 (Vol. 104, ASP Conf. Ser., Kluwer, Dordrecht, 1996).
22. Banaszekiewicz, M. & Krivov, A. V. Hyperion as a dust source in the saturnian system. *Icarus* **129**, 289–303 (1997).
23. Esposito, L. W., Brahic, A., Burns, J. A. & Marouf, E. A. in *Uranus* (eds Bergstrahl, J. T., Miner, E. A. & Matthews, M. S.) 410–465 (Univ. Arizona Press, Tucson, 1991).
24. Colwell, J. E. & Esposito, L. W. A model of dust production in the Neptune ring system. *Geophys. Res. Lett.* **17**, 1741–1744 (1990).
25. Soter, S. *The Dust Belts of Mars* (Report of Center for Radiophysics and Space Research No. 462, 1971).
26. Krivov, A. V. & Hamilton, D. P. Martian dust belts: Waiting for discovery. *Icarus* **128**, 335–353 (1997).

Acknowledgements. We thank the Galileo project at JPL for effective and successful mission operations. A.K. thanks his colleagues in the Heidelberg dust group for their hospitality and for funding his stay at MPIK. This work was supported by Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR).

Correspondence and requests for materials should be addressed to H.K. (e-mail: krueger@galileo.mpi-hd.mpg.de).

Percolative phase separation underlies colossal magnetoresistance in mixed-valent manganites

M. Uehara[†], S. Mori[‡], C. H. Chen[‡] & S.-W. Cheong[†]

^{*} Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA

[†] Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

Colossal magnetoresistance¹—an unusually large change of resistivity observed in certain materials following application of magnetic field—has been extensively researched in ferromagnetic perovskite manganites. But it remains unclear why the magnetoresistive response increases dramatically when the Curie temperature (T_C) is reduced. In these materials, T_C varies sensitively with changing chemical pressure; this can be achieved by introducing trivalent rare-earth ions of differing size into the perovskite structure^{2–4}, without affecting the valency of the Mn ions. The chemical pressure modifies local structural parameters such as the Mn–O bond distance and Mn–O–Mn bond angle, which directly influence the case of electron hopping between Mn ions (that is, the electronic bandwidth). But these effects cannot satisfactorily explain the dependence of magnetoresistance on T_C . Here we demonstrate, using electron microscopy data, that the prototypical (La,Pr,Ca)MnO₃ system is electronically phase-separated into a sub-micrometre-scale mixture of insulating regions (with a particular type of charge-ordering) and metallic, ferromagnetic domains. We find that the colossal magnetoresistive effect in low- T_C systems can be explained by percolative transport through the ferromagnetic domains; this depends sensitively on the rela-

[†] Permanent addresses: Department of Physics, Aoyama-Gakuin University, Tokyo 157-8572, Japan (M.U.); Department of Physics, Tokyo Institute of Technology, Tokyo 152-8551, Japan (S.M.).

tive spin orientation of adjacent ferromagnetic domains which can be controlled by applied magnetic fields.

In mixed-valent manganites, ferromagnetic coupling between localized Mn t_{2g} spins is mediated by the hopping of e_g electrons which enables the avoidance of the Hund's-rule energy (the so-called double-exchange mechanism)⁵. The reduction of bandwidth (t) by chemical pressure can be, therefore, consistent with the T_C decrease in the double-exchange mechanism. However, the detailed structural studies combined with band-structure considerations clearly indicate that the change in t is too small to account for the large decrease of T_C (refs 6–8). Furthermore, recent neutron-scattering experiments on low-energy magnetic excitations in manganites with various T_C show that the spin wave stiffness, proportional to t in the double-exchange mechanism, changes little in different- T_C manganites^{9,10}. On the other hand, accumulating theoretical and experimental evidence indicates that the important feature of the manganites is the competition between double-exchange ferromagnetism and another instability associated with electron–lattice coupling, partially the Jahn–Teller-type, leading to (Jahn–Teller) polaron formation^{11–14}. This polaronic view is consistent with overall trends of manganite physics, but is not successful

in explaining the enormous magnitude of magnetoresistance (MR) for low- T_C manganites^{8,12}. Here we show that the instability relevant for colossal magnetoresistance (CMR) is a static charge-ordered (CO) state with a particular modulation, resulting in the large-scale coexistence of this CO phase with a ferromagnetic (FM) metallic phase.

Earlier reports showed that T_C is optimized for x (divalent ion or carrier concentration) = $3/8$ in $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$, and also in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ (ref. 7). Therefore, the system of $\text{La}_{1-x-y}\text{Pr}_y\text{Ca}_x\text{MnO}_3$ with $x = 3/8$ was chosen to investigate the detailed chemical pressure effects. A few striking features can be seen in the temperature (T) dependence of resistivity (ρ) and MR of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}\text{MnO}_3$ shown in Fig. 1. First, T_C , determined by the maximum slope of the sudden ρ drop, decreases gradually with increasing y (for $0.25 \geq y \geq 0$), but T_{CO} (charge ordering temperature ≈ 210 K)—which is observed only for $y \geq 0.3$ and characterized by ρ upturn—does not depend on y . T_C is strongly reduced from ~ 210 K for $y = 0.25$ down to ~ 80 K for $y = 0.3$, clearly indicating the competition between the two dominant FM and CO ground states in the manganites⁷. The most striking aspect, however, is that enormously large residual ρ (for example, $\rho_0 \approx 2 \times 10^4 \Omega \text{ cm}$ for $y = 0.4$), much

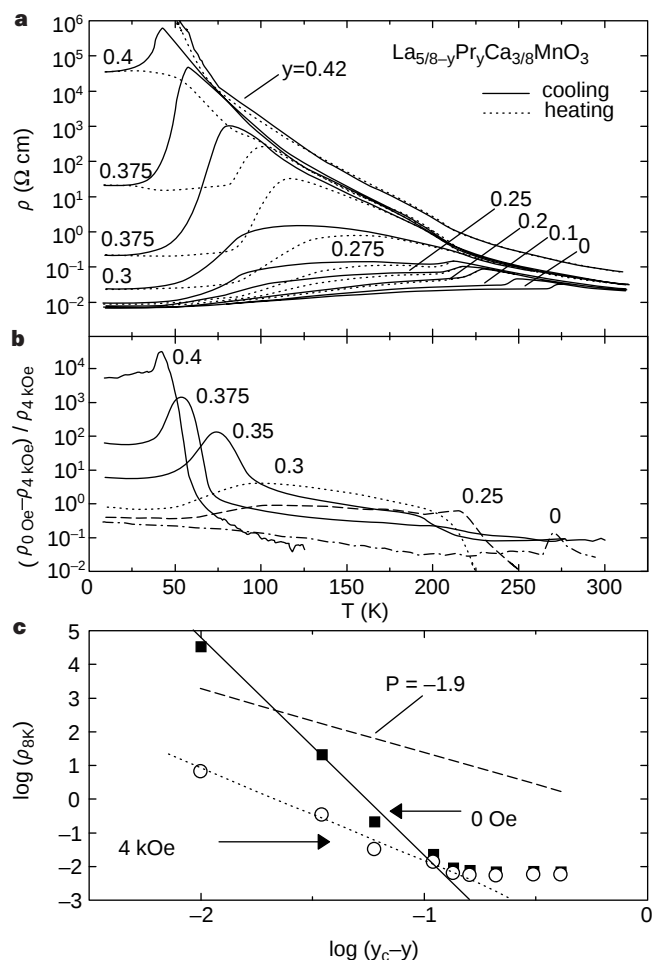


Figure 1 Transport and magnetic properties of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}\text{MnO}_3$ as a function of temperature and y . **a**, T dependence of ρ . Both cooling (solid lines) and heating (dotted lines) curves are shown. **b**, Magnetoresistance of representative specimens in 4 kOe with field cooling. **c**, $\log \rho_0$ versus $\log (y_c - y)$ plot for zero field (filled squares) and 4 kOe (open circles) with $y_c = 0.41$. For $0.4 \geq x \geq 0.275$, there exists a linear relationship with the slope of -6.9 for zero field (-2.6 for 4 kOe). The dashed line with the slope of -1.9 represents the prediction of percolation theory.

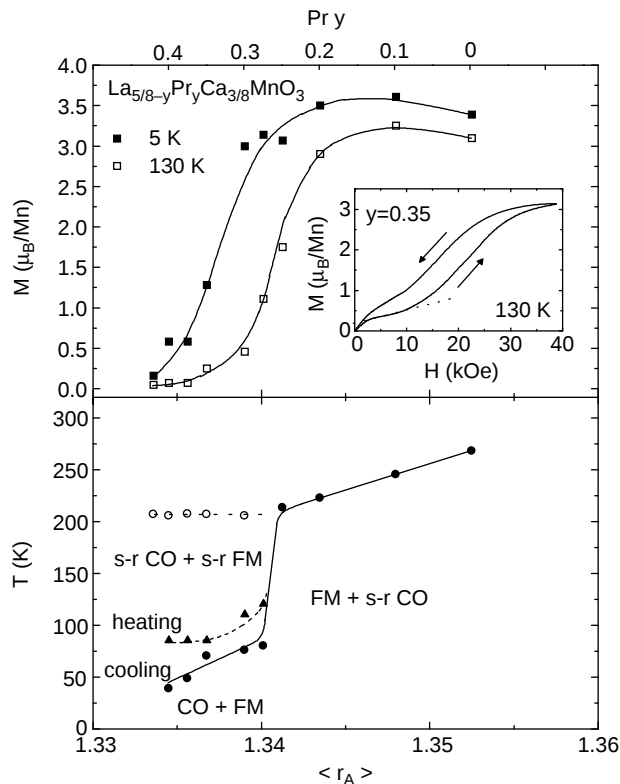


Figure 2 Ferromagnetically-ordered moment and the phase diagram of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}\text{MnO}_3$ as a function of average ionic radius ($\langle r_A \rangle$) or y . Top panel, saturation moments at 5 K (filled squares) and 130 K (open squares) of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}\text{MnO}_3$ as a function of y of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}$. Solid lines are guides for the eye. Inset, $M(H)$ for $y = 0.35$ at 130 K. The FM component is evident as indicated with a dashed line, and there exists a field-induced transition (the melting of CO state) in high fields. Bottom panel, the phase diagram of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}\text{MnO}_3$ as a function of $\langle r_A \rangle$ of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}$. T_C and T_{CO} are shown as filled circles (or triangles) and open circles, respectively. There exists a large thermal hysteresis at T_C for $y \geq 0.275$. The coexistence of l-r FM/s-r CO, s-r CO/s-r FM, or l-r CO/l-r FM is schematically shown in the phase diagram (l-r indicates long range, s-r short range). Note that the average ionic radius, $\langle r_A \rangle$, was estimated from the tabulated values for 12 coordination³³.

larger than the Mott metallic limit, develops for large y even though the T dependence of ρ is metallic-like below T_C , suggesting percolative conduction through metallic regions embedded in insulating regions. A significant hysteresis develops when T_C becomes lower than $T_{co} \approx 210$ K (Fig. 1), consistent with the first-order-type transition in a system with two-phase coexistence. Even for $0.275 \geq y \geq 0.2$ where no CO transition is indicated in $\rho(T)$, the ρ hysteresis is clearly observable, and this hysteresis behaviour appears to correlate with a broad 'hump' far below T_C which exists for y down to 0.

We have measured the MR of $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}\text{MnO}_3$ (Fig. 1b) in a magnetic field of 4 kOe, which is high enough to orient magnetic domains but low enough not to affect the CO phase¹⁵. The MR peaks near T_C result from the shift of T_C to higher T by applying field, and the MR becomes "colossal" in low- T_C materials (particularly for $0.4 \geq y \geq 0.25$). One noticeable feature is the existence of a plateau in the MR for $0.375 \geq y \geq 0.25$ at ~ 210 K $\geq T \geq \sim 80$ K; this MR plateau is also associated with the two-phase coexistence. The FM behaviour at 130 K, between T_{co} and T_C , for $y = 0.35$ is clearly visible in the magnetization, $M(H)$, curve (Fig. 2 inset), and the change of this FM component at 130 K with y is shown in the top panel of Fig. 2. The persistence of the FM component at 130 K for y up to 0.42 indicates the two-phase coexistence even for $T_{co} \geq T \geq T_C$. The saturation moment (M_s) at 5 K in the same panel decreases sharply with y (≥ 0.3) where T_C is suppressed below ~ 80 K, supporting the suggestion above that two phases coexist for $y \geq 0.3$ below T_C . We note that M_s for $0.3 > y \geq 0$, including $y = 0$, at 5 K is also slightly reduced from the optimum value of $\sim 3.9\text{--}4\mu_B$ in $(\text{La}_{1-x}\text{Sr}_x)\text{MnO}_3$ ($x \approx 3/8$; $T_C \approx 375$ K). This reduced M_s suggests the presence of a small amount of CO phase even for $y = 0$.

Direct evidence of the two-phase coexistence is provided by our electron microscopy study. Variations of the abundance and size of CO domains with y and T were investigated by electron diffraction and dark-field imaging experiments. The dark-field image, obtained from a superlattice peak due to the CO phase at 20 K for $y = 0.375$ (Fig. 3a), clearly shows the existence of CO domains below T_C . In Fig. 3a, the white region is a CO domain, and the highlighted dark region is a charge-disordered domain which is presumably FM. Much to our surprise, the typical size of CO and FM domains at low T is of the order of several thousand ångströms. This sub-micrometre-scale mixture is commonly observed in other compositions below T_C , and this is in sharp contrast with the nanoscale mixture of CO and FM phases in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ at ~ 220 K $> T > \sim 150$ K (refs 16, 17). We found that the volume fraction of CO domains at 20 K increases rapidly with y , which is consistent with the dramatic increase of ρ_0 with y as shown in Fig. 1. We note the presence of superlattice peaks characteristic of $x = 1/2$ in all our specimens with $x = 3/8$ (that is, the modulation wavelength is $2a$ with $a \approx 5.5$ Å)¹⁸. In spite of the presence of this $x = 1/2$ -type CO state, the carrier concentration in the CO domains is presumably close to $3/8$, because it must be difficult to break the charge neutrality in the large length scale of thousands of ångströms. This difference in carrier concentrations could be accommodated by charge defects in the $x = 1/2$ -type CO state. In a typical chemical phase separation such as spinodal decomposition, ionic diffusion length fixes the length scale for chemical inhomogeneity¹⁹. The relatively large length scale (~ 0.5 µm) in our phase separation at low T may originate from the large electronic diffusion coefficient in the electronically inhomogeneous system.

We have also followed the microstructural changes of the CO domains as T is raised. Figure 3b and c shows the dark field images obtained from the same area (for $y = 0.4$) at 17 K and 120 K, respectively. The sub-micrometre-scale CO regions with relatively uniform contrast at 17 K show a fine, nanoscale speckled contrast at 120 K, indicating the coexistence of CO (bright speckles) and charge-disordered (dark speckles) nanodomains. Considering the non-zero saturation moment at 130 K for y up to 0.42, the charge-

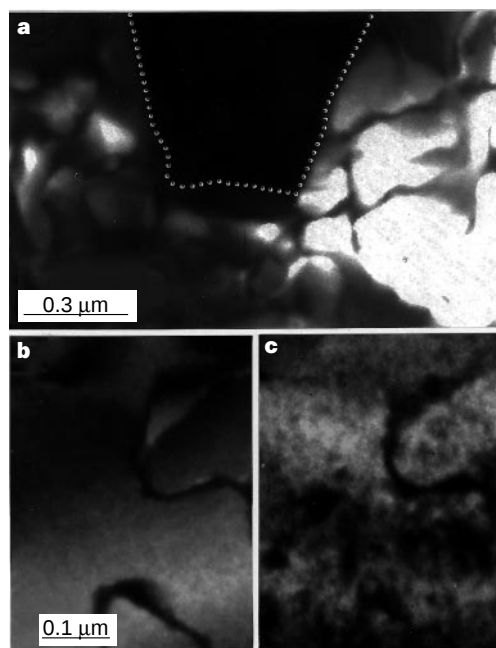


Figure 3 Dark-field images for $\text{La}_{5/8-y}\text{Pr}_y\text{Ca}_{3/8}\text{MnO}_3$ obtained by using a superlattice peak caused by CO. Panel **a** shows the coexistence of charge-ordered (insulating) and charge-disordered (FM metallic) domains at 20 K for $y = 0.375$. The charge-disordered domain (dark area) is highlighted with dotted lines for clarity. The curved dark lines present in CO regions are antiphase boundaries, frequently observed in dark-field images for the commensurate CO states of $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ (ref. 18). Panels **b** and **c**, obtained from the same area for $y = 0.4$ at 17 K and 120 K, respectively, show the development of nanoscale charge-disordered domains at $T > T_C$. The curved lines in **a**, **b** and **c** signify the presence of antiphase boundaries of the CO domains.

disordered nanodomains are believed to be FM. The nanoscale existence of the CO and FM domains at finite temperatures, which is similar to that observed in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$, is also observed in other compositions of $0.4 > y > 0.25$.

We can now establish a clear two-phase picture, as shown schematically in Fig. 4, to understand the large ρ_0 and the large MR in reduced- T_C manganites. We first note that the ratio of the saturation moment ($\sim 0.6\mu_B$) for $y = 0.4$ (which is close to the critical concentration (y_c) for the metal–insulator transition) and the full moment of $\sim 4\mu_B$ is within the three-dimensional percolation threshold range of 10–25% (ref. 20). To compare the behaviour of $\log(y_c - y)$ with $y_c \approx 0.41$ (Fig. 1c). There exists a linear region ($0.4 \geq y \geq 0.275$) in this plot where ρ_0 changes by about 7 orders of magnitude. The slope ($P \approx -6.9$) of the linear region turns out to be much larger than the three-dimensional percolation prediction²⁰ (~ -1.9). Electron conduction between neighbouring FM domains should be reduced if their magnetizations are not aligned (Fig. 4a). This conduction reduction must be particularly large in manganites where hopping electrons are strongly aligned with the local orientation of magnetization (that is, manganites are the so-called half-metallic ferromagnets)²¹. The degree of this magnetization misalignment should increase when the concentration of Pr approaches the percolation threshold where the abundance of CO domains leads to poor connection of the minority FM domains. Therefore, with decreasing relative volume of FM domains, ρ_0 increases much faster than the prediction of simpler percolative transport. If all these magnetizations of FM domains were aligned in applied fields (Fig. 4b), ρ_0 would follow closely the percolation prediction. To

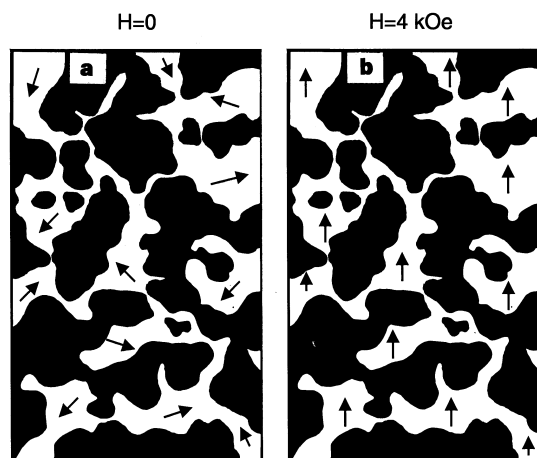


Figure 4 Schematic illustration of the sub-micrometre-scale coexistence of the $x = 1/2$ -type CO insulating (dark area) and FM metallic (white area) domains. The typical size of domains is $\sim 0.5 \mu\text{m}$. In zero field (**a**), the magnetizations of FM domains are random, but all magnetizations of FM domains can be aligned by applying field of about 4 kOe (**b**). With the variation of y (that is, the abundance of CO domains), the residual ρ_0 follows $\rho_0 \propto (y_c - y)^P$ with $P \approx -6.9$ and $P \approx -2.6$ **a** and **b**, respectively.

verify this idea, we show the variation of ρ_0 (open circles) with y (measured in a field of 4 kOe) in Fig. 1c. The slope of the linear region is now about -2.6 , only slightly different from the percolation prediction of ~ -1.9 . This slight difference may originate from the proximity of FM regions to CO regions. For example, electron scattering at the CO/FM interfaces or strains from adjacent CO regions could influence the resistivity of FM regions. Thus, the change of ρ_0 from $P \approx -6.9$ to $P \approx -2.6$ behaviour with magnetic domain rotation induces CMR.

All of our findings can be summarized within the two-phase picture. (1) When T_C is above ~ 210 K, but less than the optimum value of ~ 375 K, there exists a long-range FM state with a short-range CO phase. (2) When the upturn of ρ at T_{CO} is clearly visible (for $0.4 \geq y \geq 0.3$), T_C is reduced to a value below ~ 80 K, short-range CO and short-range FM coexist at $T_{CO} \geq T \geq T_C$, and long-range FM with long-range CO develops below T_C . This view, as well as the experimental values of T_C and T_{CO} as a function of Pr concentration, is summarized in the bottom panel of Fig. 2.

Our results show that the phase separation relevant to CMR is the one between the $x = 1/2$ -type CO and FM states with sub-micrometre-scale lengths, which is *not* a charge-segregation type of phase separation. Doped carriers tend to cluster, forming FM domains, when charge carriers are introduced into the anti-ferromagnetic, insulating LaMnO_3 (refs 14, 22–30). This carrier clustering, that is, charge segregation, can occur only within small length scales because of charge neutrality, and the large length scale ($\sim 0.5 \mu\text{m}$) of our phase separation clearly shows that the phase separation is a new type, not driven by the charge segregation. As we discussed earlier, the insulating phase after phase separation in our specimens is always the $x = 1/2$ -type CO state. The peculiarity of this phase can be seen from the fact that a magnetic field readily converts it to a FM metallic state, but the insulating state of LaMnO_3 (even with low carrier doping) is insensitive to magnetic field^{7,15}. This coexistence is not the direct consequence of simple chemical inhomogeneity. For example, electron microprobe analysis and X-ray diffraction experiments do not show any indication of two phases with different chemical composition. Furthermore, simple chemical phase separation is inconsistent with the systematic

decrease of T_C with Pr doping. However, a slight fluctuation of chemical compositions could nucleate one electronic phase over the other electronic phase.

Our experimental findings provide a coherent picture for the earlier experimental results and ambiguities. FM (metallic) and the $x = 1/2$ -type CO states coexist in a broad range of phase space, and the percolative transport through this sub-micrometre-scale two-phase mixture is responsible for the anomalous behaviour of low- T_C manganites. The subtle balance between these two states with distinctly different electronic properties can be readily influenced by varying physical parameters (such as the intensity of an applied magnetic field, the chemical pressure, and oxygen-isotope composition), producing various ‘‘colossal’’ effects^{1–4,28–32}. □

Received 13 January; accepted 30 March 1999.

- Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La-Ca-Mn-O films. *Science* **264**, 413–415 (1994).
- Hwang, H. Y. *et al.* Lattice effects on the magnetoresistance in doped LaMnO_3 . *Phys. Rev. Lett.* **75**, 914–917 (1995).
- Coey, J. M. D. *et al.* Electronic localization in mixed-valence manganites. *Phys. Rev. Lett.* **75**, 3910–3913 (1995).
- Khazeni, K. *et al.* Effect of pressure on the magnetoresistance of single crystal $\text{Nd}_{0.5}\text{Sr}_{0.36}\text{Pb}_{0.14}\text{MnO}_{3-\delta}$. *Phys. Rev. Lett.* **76**, 295–298 (1996).
- Zener, C. Interaction between the d-shells in the transition metals II. Ferromagnetic compounds of manganese with perovskite structure. *Phys. Rev.* **82**, 403–405 (1951).
- Radaelli, P. G. *et al.* Structural effects on the magnetic and transport properties of perovskite $\text{A}_{1-x}\text{A}'_x\text{MnO}_3$. *Phys. Rev. B* **56**, 8265–8276 (1997).
- Cheong, S.-W. & Hwang, H. Y. In *Colossal Magnetoresistance Oxides* (ed. Tokura, Y.) Ch. 7 (Monogr. in Condensed Matter Sci., Gordon & Breach, London, 1999).
- Laukhin, V. *et al.* Pressure effects on the metal-insulator transition in magnetoresistive manganese perovskites. *Phys. Rev. B* **56**, R10009–R10012 (1997).
- Hwang, H. Y. *et al.* Softening and broadening of the zone boundary magnons in $\text{Pr}_{0.65}\text{Sr}_{0.37}\text{MnO}_3$. *Phys. Rev. Lett.* **80**, 1316–1319 (1998).
- Fernandez-Baca, J. A. *et al.* Evolution of the low-frequency spin dynamics in ferromagnetic manganites. *Phys. Rev. Lett.* **80**, 4012–4015 (1998).
- Millis, A. J., Littlewood, P. D. & Shraiman, B. I. Double exchange alone does not explain the resistivity of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **74**, 5144–5147 (1995).
- Millis, A. J., Shraiman, B. I. & Mueller, R. Dynamic Jahn-Teller effect and colossal magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 175–178 (1996).
- Roder, H., Zang, J. & Bishop, A. R. Lattice effects in the colossal-magnetoresistance manganites. *Phys. Rev. Lett.* **76**, 1356–1359 (1996).
- Zhou, J.-S. & Goodenough, J. B. Phonon-assisted double exchange in perovskite manganites. *Phys. Rev. Lett.* **80**, 2665–2668 (1998).
- Tomioka, Y. *et al.* Magnetic-field-induced metal-insulator phenomena in $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ with controlled charge-ordering instability. *Phys. Rev. B* **53**, R1689–R1692 (1996).
- Chen, C. H. & Cheong, S.-W. Commensurate to incommensurate charge ordering and its real-space images in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$. *Phys. Rev. Lett.* **76**, 4042–4045 (1996).
- Mori, S., Chen, C. H. & Cheong, S.-W. Paired and unpaired charge stripes in the ferromagnetic phase of $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$. *Phys. Rev. Lett.* **81**, 3972–3975 (1998).
- Chen, C. H., Cheong, S.-W. & Hwang, H. Y. Charge-ordered stripes in $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ with $x > 0.5$. *J. Appl. Phys.* **81**, 4326–4330 (1997).
- Hilliard, J. E. In *Phase Transformations* (ed. Cohen, M.) Ch. 12 (Am. Soc. for Metals, Metals Park, Ohio, 1970).
- Shklovskii, B. I. & Efros, A. L. *Electronic Properties of Doped Semiconductors* (Springer, New York, 1984).
- Hwang, H. Y. & Cheong, S.-W. In *Colossal Magnetoresistance Oxides* (ed. Tokura, Y.) Ch. 9 (Monogr. in Condensed Matter Sci., Gordon & Breach, London, 1999).
- Emery, V. J. & Kivelson, S. A. Frustrated phase separation and high-temperature superconductivity. *Physica C* **209**, 597–621 (1993).
- Nagaev, E. L. Phase separation in degenerate semiconductors and high-temperature superconductors. *Phys. Status Solidi B* **186**, 9–42 (1994).
- Goodenough, J. B. & Zhou, J.-S. New forms of phase segregation. *Nature* **386**, 229–230 (1997).
- De Teresa, J. M. *et al.* Evidence for magnetic polarons in the magnetoresistive perovskites. *Nature* **386**, 256–259 (1997).
- Yunoki, S. *et al.* Phase separation in electronic models for manganites. *Phys. Rev. Lett.* **80**, 845–848 (1998).
- Allodi, G. *et al.* Electronic phase separation in lanthanum manganites: evidence from ^{55}Mn NMR. *Phys. Rev. B* **56**, 6036–6046 (1997).
- Kiryukhin, V. *et al.* An X-ray induced insulator–metal transition in a magnetoresistive manganite. *Nature* **386**, 813–815 (1997).
- Asamitsu, A. *et al.* Current switching of resistive states in magnetoresistive manganites. *Nature* **388**, 50–52 (1997).
- Fiebig, M. *et al.* Visualization of the local insulator-metal transition in $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$. *Science* **280**, 1925–1928 (1998).
- Rodríguez-Martínez, L. M. & Attfield, J. P. Cation disorder and size effects in magnetoresistive manganese oxide perovskites. *Phys. Rev. B* **54**, R15622–R15625 (1996).
- Zhao, G. *et al.* Giant oxygen isotope shift in the magnetoresistive perovskite $\text{La}_{1-x}\text{Ca}_x\text{MnO}_{3-y}$. *Nature* **381**, 676–678 (1996).
- Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **32**, 751–767 (1976).

Acknowledgements. We benefited from the transport system set up by C. Hess, and we thank M. Gershenson, A. Ruckenstein, G. Kotliar, A. J. Millis and N. Furukawa for discussions. We also thank B. Batlogg for magnetic susceptibility measurements. M.U. and S.-W.C. were supported by the NSF, and M.U. acknowledges support from the JPSJ Fellowships for Young Scientists.

Correspondence and requests for materials should be addressed to S.-W.C. (e-mail: sangc@physics.rutgers.edu).

Dynamics of individual flexible polymers in a shear flow

Philip LeDuc*, Charbel Haber†, Gang Bao* & Denis Wirtz†‡

* Department of Mechanical Engineering, † Department of Chemical Engineering, ‡ Department of Materials Science and Engineering, The Johns Hopkins University, Baltimore, Maryland 21218, USA

Polymer dynamics are of central importance in materials science, mechanical engineering, biology and medicine^{1,2}. The dynamics of macromolecular solutions and melts in shear flow are typically studied using bulk experimental methods such as light and neutron scattering and birefringence^{3,4}. But the effect of shear on the conformation and dynamics of individual polymers is still not well understood^{5–7}. Here we describe observations of the real-time dynamics of individual, flexible polymers (fluorescently labelled DNA molecules^{8–15}) under a shear flow. The sheared polymers exhibit many types of extended conformation with an overall orientation ranging from parallel to perpendicular with respect to the flow direction. For shear rates much smaller than the inverse of the relaxation time of the molecule, the relative populations of these two main types of conformation are controlled by the rate of the shear flow. These results question the adequacy of assumptions made in standard models of polymer dynamics^{5,6}.

To visualize the deformation of individual flexible polymers, we use a dilute solution (20 ng ml⁻¹, much smaller than the overlap concentration of ~0.13 mg ml⁻¹) of DNA from the T2 phage that infects *Escherichia coli* suspended in Tris-EDTA buffer and an oxygen-scavenging system to reduce photobleaching. In these conditions, DNA is a flexible polymer with a contour length of ~56 µm, a persistence length of 0.12 µm, a quiescent radius of gyration of 1.5 µm, and an equilibrium relaxation time of $\tau = 1.0 \pm 0.2$ s (measured via dynamic light scattering). DNA molecules are stained with a dye (YOYO-1, Molecular Probes, Eugene, Oregon) for one hour and then placed on a fixed coverslip¹⁶, which is attached to a computer-controlled sliding stage and centred

under an upright microscope (Fig. 1a). The coverslip and sliding stage form the upper and lower plates of the shear flow-cell. This apparatus generates a uniform, laminar shear flow.

We use an SIT (silicon-intensifier target) camera along with an image processor (Hamamatsu, Bridgewater, New Jersey) to capture two-dimensional images, which are recorded in real-time onto S-VHS video and subsequently analysed using a frame grabber. The plane of the two-dimensional images is parallel to the plane formed by the velocity direction, $\mathbf{v}$ ($v = |\mathbf{v}|$), and the vorticity direction, $\mathbf{w}$, and thus perpendicular to the velocity gradient, $\nabla\mathbf{v}$ (see Fig. 1b). These images are categorized according to the conformational shape displayed by each individual DNA molecule. The entire system is monitored and controlled through a data acquisition system which produces a closed-loop feedback system providing a spatial resolution of motion smaller than one micrometre. Possible wall effects (long-range hydrodynamic interactions between the polymer and the walls of the flow-cell) on molecular conformation are alleviated by using a gap space of 100 µm, which is much larger than the radius of gyration of the molecules (similar results were obtained for gap widths ranging from 100 to 300 µm). Molecules are monitored at the mid-plane of the flow-cell (wall effects appear only when near the walls), with 1-µm depth of focus. The uniformity of the flow field was checked using 1.94-µm-radius polystyrene fluorescent spheres (Duke Scientific Corp., Palo Alto, California) subjected to

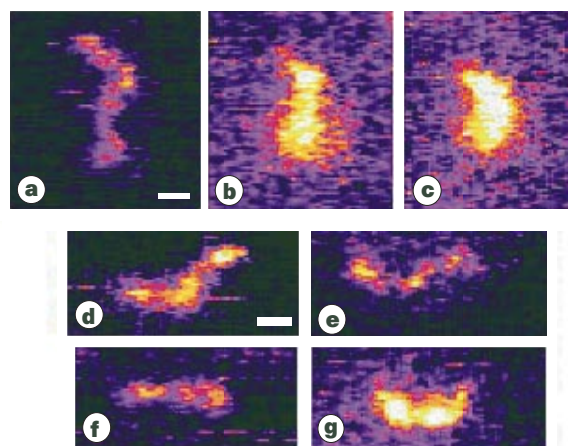
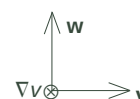


Figure 2 The different conformations of DNA under shear. These can be categorized into two main conformational classes: perpendicular or parallel to the flow direction. Although DNA under shear can be loosely categorized into seven subtypes of conformations, **a–g**, we find that these molecules adopt a continuous spectrum of configurations. **a–g**, Images of fluorescently labelled DNA molecules; **a**, conformation characterized by the long thin nature of the DNA and an alignment perpendicular to the flow direction; **b**, conformation characterized by a shorter length and larger width of the DNA than **a**, and alignment in the direction perpendicular to the shear flow; **c**, conformation characterized by the ends of the DNA being in front or behind the main fluorescence region of the DNA molecule, as well as by alignment perpendicular to the direction of shear flow; **d**, conformation characterized by its long thin body and its end-to-end alignment at an angle between 30° and 60° to the direction of the shear flow; **e**, conformation characterized by a long thin body and an alignment parallel to the flow direction; **f**, conformation characterized by a shorter length and larger width of the DNA compared to **e** and its alignment parallel to the direction of the shear flow; **g**, conformation characterized by the ends of the DNA being above or below the main fluorescence region of the DNA molecule and its alignment parallel to the flow direction. Scale bar in **a** is 1 µm and applies to **a–g**. Shear rate is 0.10 s⁻¹.

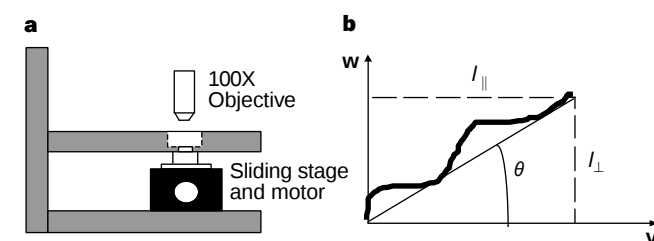


Figure 1 Shear-flow experimental set-up and sketch of an individual DNA molecule. **a**, Diagram of the experimental set-up used to image individual DNA molecules at small strain rates. The top coverslip is immobilized by attaching it to a 2-axis stage. This stage is connected to a Uni-Slide continuous drive system in order to alleviate out-of-plane motion and to a linear variable differential transformer in order to measure the stage displacements. This system is monitored and controlled by a closed-loop computer-controlled system. **b**, For every DNA molecule observed, four quantities are calculated to determine the statistical variation at different strain rates. The contour length is obtained by measuring the total distance along the fluorescent region of the DNA. The parallel and perpendicular lengths are the absolute lengths of the end-to-end segment of the molecule projected onto the axis parallel and perpendicular to the direction of shear flow, respectively. The angle is calculated from the ends of the molecule when compared to the direction of the shear flow. Two-dimensional images of the polymers are recorded in the plane formed by the velocity direction, $\mathbf{v}$, and the vorticity direction, $\mathbf{w}$, and thus perpendicular to the velocity gradient, $\nabla\mathbf{v}$, of the shear flow.

the same shear flow conditions as the DNA molecules used in the experiments. As expected, the flow generated by the parallel plates is uniform (data not shown)¹⁷.

After systematic observations of individual DNA molecules ($n = 210$) under well defined strain rates, $\dot{\gamma}$, we find that sheared DNA molecules adopt two main types of conformation: either parallel or perpendicular to the flow direction. Relative orientations are calculated from the angle created from the ends of the DNA segments when compared to the direction of flow, and shapes are determined by regions of relative fluorescence intensity along the DNA molecule (Fig. 2; polymers which are out of focus are not considered). Although DNA under shear can be loosely categorized into seven subtypes of conformation (Fig. 2), we find that these molecules adopt a continuous spectrum of configurations. Extended conformations are seemingly absent under no-shear conditions as quiescent DNA molecules display featureless, circular fluorescent regions (data not shown; see also ref. 18).

We observe that a sheared DNA molecule can change both size and orientation dynamically. In Fig. 3a we show a DNA molecule, initially oriented in the direction perpendicular to the flow direc-

tion, undergoing momentary changes of conformation while maintaining an overall vertical orientation. The extended conformations of sheared DNA molecules typically survive for timescales longer than the relaxation time of an individual molecule in the quiescent state (that is, >1.0 s). The average time that it takes for a sheared DNA molecule to switch from one of the two main conformational classes (that is, with an overall orientation either parallel or perpendicular to the flow direction) to the other is 4.6 s, 3.9 s, and 3.3 s at shear rates of 0.05 s^{-1} , 0.10 s^{-1} , and 0.20 s^{-1} , respectively. It is remarkable that polymers, which are in close proximity to one another, can undergo dramatically different conformational changes. In Fig. 3b we show two DNA molecules; one changes its conformational changes with $>90^\circ$ rotation.

We further quantified the conformation of individual polymers under shear by computing the instantaneous and time- (or ensemble-) averaged contour length of the DNA molecules, l and $\langle l \rangle$, the length of the projections of the molecule in the directions parallel and perpendicular to the flow direction, $l_\perp$, $\langle l_\perp \rangle$, $l_\parallel$, $\langle l_\parallel \rangle$, and the angle of the molecule with respect to the flow direction, θ and $\langle \theta \rangle$, respectively (see Fig. 1b for definitions). We find that the instantaneous molecular orientation parameters, $l_\perp$, $l_\parallel$ and θ , vary in a seemingly random fashion (Fig. 4). However, in agreement with well-documented bulk measurements of the same quantities, sheared polymers have a tendency to align in the direction parallel to that of the shear flow (Fig. 4). Ensemble-averaged (Fig. 4a and b) and time-averaged (Fig. 4c and d) angles of orientation are $\langle \theta \rangle \approx 0^\circ$ at all tested shear rates. Similarly, $\langle l_\perp \rangle$ decreases, and $\langle l_\parallel \rangle$ increases, for increasing shear rates (Fig. 4d). The population of the two main conformational classes is influenced by the shear rate of the flow. As

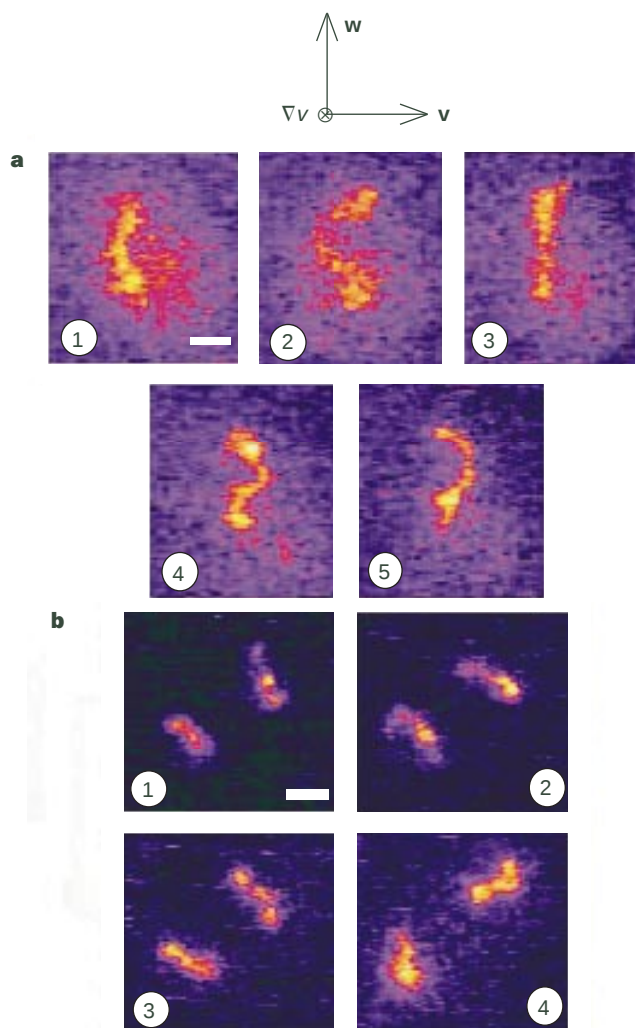


Figure 3 Dynamic motion of DNA molecules experiencing well defined shear strains. **a**, Time-lapse snapshots of a fluorescently labelled DNA molecule subjected to a shear rate of 0.2 s^{-1} . Scale bar in 1 is $1.5 \mu\text{m}$ and applies to 1–5; these numbers indicate the order of the images. The time between images is 2 s. See text for comments. **b**, Time-lapse snapshots of two fluorescently labeled DNA molecules subjected to a shear rate of 0.10 s^{-1} , separated from each other by $\sim 6 \mu\text{m}$. See text for comments. The time between each frame is 2 s. Scale bar in 1 is $2 \mu\text{m}$ and applies to 1–4.

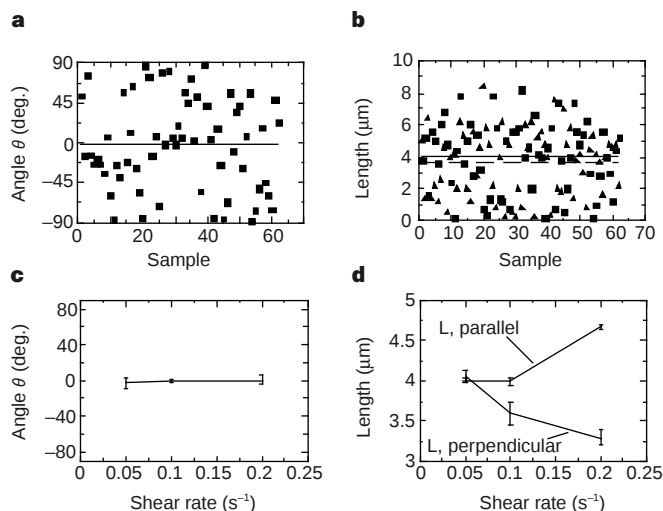


Figure 4 Quantitative characterization of the conformation of individual DNA molecules under shear. **a**, Instantaneous variation (filled squares) and ensemble-averaged value (line) of the angle between sheared polymers and the shear flow direction for a sequential set of samples at the same shear rate. **b**, Instantaneous variation (filled squares, filled triangles) and ensemble-averaged value (line, dashed line) of respectively parallel and perpendicular absolute lengths for a sequential set of images. For **a** and **b**, samples are taken sequentially and monitored while in the field of vision at a shear rate of 0.20 s^{-1} . **c**, Time-averaged angle of orientation of sheared DNA molecules as a function of shear rate. The ensemble-averaged angle of orientation (line) is around zero for all of the shear strain rates tested. **d**, Time-averaged parallel and perpendicular lengths, $\langle l_\parallel \rangle$ and $\langle l_\perp \rangle$, as a function of shear rate. The relative populations of the two main conformational classes for different shear rates (respectively 0.05 , 0.1 , 0.2 s^{-1}) are as follows. Vertical conformation (per cent): 42, 35, 23; horizontal conformation (per cent): 36, 44, 58. Percentages are based on the total number of DNA molecules that are in focus and the qualitative categorization by shape and orientation of the molecules. Molecules that are out of the focal plane of the fluorescence microscope are not considered.

the shear rate increases, the DNA molecules tend to align more towards the flow direction (see Fig. 4 legend).

Polymers extended in the direction of the shear flow have been qualitatively predicted by classical theories, and have been indirectly observed by using bulk birefringence^{3,4}. Partial stretching of individual polymers in the direction parallel to the flow direction is due to the viscous drag exerted on the molecule⁵, which has been invoked to explain shear-thinning in polymer solutions¹. However, the existence of vertical conformations (such as those shown in Fig. 2a–c) is unexpected, and is not considered by classical models of polymer physics^{5,6}. Classically, it is assumed that, in the plane parallel to the velocity and the velocity-gradient directions, shear orients polymers at an angle $\leq 45^\circ$ with the flow direction. Shear may also induce tumbling of the polymers within such a plane. However, in the plane parallel to the velocity and the vorticity directions (the present plane of visualization; see Fig. 1), neither shear-induced alignment nor polymer tumbling can explain the existence of vertical conformations. Although fluid-mechanics models—such as Jeffery orbits describing the motion of axisymmetric rigid particles in a shear flow—may serve as limiting-case analyses^{19–23}, they cannot be directly applied to the dynamics of flexible polymers reported here. Furthermore, classical polymer theories predict⁵ that a flexible polymer becomes stretched and oriented when the Weissenberg number, $We = \dot{\gamma}\tau$, is equal to or larger than unity. But here we find shear-induced deformation of flexible polymers can occur at Weissenberg numbers much smaller than unity (Figs 1–4).

Our experimental results show that conventional approaches, such as birefringence and light scattering, which measure only ensemble-averaged molecular parameters (that is, $\langle \theta \rangle$, $\langle l \rangle$), overlook the extremely rich dynamics of individual polymers under shear. Our observations also provide an insight into the conformational and orientational changes of polymers in a shear flow, and a basis for further theoretical modelling. The molecular-level approach to non-equilibrium polymer physics that we describe here could be readily extended to many other polymer systems, including entangled solutions of flexible polymers and semiflexible polymers such as actin^{24,25}. □

Received 21 October 1998; accepted 22 March 1999.

1. Munk, P. *Introduction to Macromolecular Science* (Wiley & Sons, New York, 1989).
2. Hoffman, A., Ratner, B. & Horbett, T. *Polymers as Biomaterials* (Plenum, New York, 1995).
3. Fuller, G. G. *Optical Rheometry of Complex Fluids* (Oxford Univ. Press, New York, 1995).
4. Janeschitz-Kriegl, H. *Polymer Melt Rheology and Flow Birefringence* (Springer, New York, 1983).
5. Doi, M. & Edwards, S. F. *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
6. de Gennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, 1991).
7. de Gennes, P. G. Molecular individualism. *Science* **276**, 199 (1997).
8. Matsumoto, M. et al. Direct observation of Brownian-motion of macromolecules by fluorescence microscope. *J. Polym. Sci.* **30**, 779–783 (1992).
9. Perkins, T. T., Quake, S. R., Douglas, D. E. & Chu, S. Relaxation of a single DNA molecule observed by optical microscopy. *Science* **264**, 822–826 (1994).
10. Perkins, T. T., Smith, D. E. & Chu, S. Single polymer dynamics in an elongational flow. *Science* **276**, 2016–2021 (1997).
11. Schwartz, D. C. & Koval, M. Conformational dynamics of individual DNA molecules during gel electrophoresis. *Nature* **338**, 520–522 (1989).
12. Shivashankar, G. & Libchaber, A. Single DNA molecule grafting and manipulation using a combined atomic force microscope and an optical tweezer. *Appl. Phys. Lett.* **7**, 3727–3729 (1997).
13. Smith, D. E., Perkins, T. T. & Chu, S. Dynamical scaling of DNA diffusion coefficients. *Macromolecules* **29**, 1372–1373 (1996).
14. Smith, S. B., Finzi, L. & Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science* **258**, 1122–1126 (1992).
15. Wirtz, D. Direct measurement of the transport properties of a single DNA molecule. *Phys. Rev. Lett.* **75**, 2436–2439 (1995).
16. Morikawa, K. & Yanagida, M. Visualization of individual DNA molecules in solution by light microscopy: DAPI staining method. *J. Biochem.* **89**, 693–696 (1981).
17. Archer, L. A., Larson, R. G. & Chen, Y. L. Direct measurement of slip in sheared polymer solutions. *J. Fluid Mech.* **301**, 133–151 (1995).
18. Matsumoto, S., Morikawa, K. & Yanagida, M. Light microscopic structure of DNA in solution studied by the 4',6-diamidino-2-phenylindole staining method. *J. Mol. Biol.* **152**, 501–516 (1981).
19. Jeffery, G. B. The motion of ellipsoidal particles immersed in a viscous fluid. *Proc. R. Soc. Lond. A* **102**, 161–179 (1922).
20. Leal, L. G. & Hinch, E. J. The effect of weak Brownian rotations on particles in shear flow. *J. Fluid Mech.* **46**, 685–703 (1971).
21. Hinch, E. J. The deformation of a nearly straight thread in a shearing flow with weak Brownian motions. *J. Fluid Mech.* **75**, 765–775 (1976).
22. Hinch, E. J. The distortion of a flexible inextensible thread in a shearing flow. *J. Fluid Mech.* **74**, 317–333 (1976).
23. Hinch, E. J. Mechanical models of dilute polymer solutions in strong flows. *Phys. Fluids* **20**, S22–S30 (1977).

24. Xu, J., Palmer, A. & Wirtz, D. Rheology and microrheology of semiflexible polymer solutions: actin filament networks. *Macromolecules* **31**, 6486–6492 (1998).
25. Ostap, E. M., Yanagida, T. & Thomas, D. D. Orientational distribution of spin-labeled actin oriented by flow. *Biophys. J.* **63**, 966–975 (1992).

Acknowledgements. We thank M. Ferro, L. Archer, J. Harden, D. Lavan, and J. van Zanten for discussions. This work was supported by NASA (D.W.), ACS-PRF (D.W.), NSF (D.W.), the Whitaker Foundation (D.W.), Merck, Inc. (D.W.), ARO (G.B.), and IMRE (G.B.).

Correspondence and requests for materials should be addressed to D.W. (e-mail: wirtz@jhu.edu).

Tuning bilayer twist using chiral counterions

R. Oda^{*†}, I. Huc^{*}, M. Schmutz[‡], S. J. Candau[†] & F. C. MacKintosh^{§†}

^{*} Institut Européen de Chimie et Biologie, ENSCPB Av. Pey Berland, BP 108, 33402 Talence Cedex, France

[†] Laboratoire de Dynamique des Fluides Complexes, 3 Rue de l'Université, 67000 Strasbourg, France

[‡] Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM/CNRS/ULP, BP 163, 67404 Illkirch Cedex, France

[§] Department of Physics, University of Michigan, Ann Arbor, Michigan 48109-1120, USA

From seashells to DNA, chirality is expressed at every level of biological structures. In self-assembled structures it may emerge cooperatively from chirality at the molecular scale. Amphiphilic molecules, for example, can form a variety of aggregates and mesophases that express the chirality of their constituent molecules at a supramolecular scale of micrometres (refs 1–3). Quantitative prediction of the large-scale chirality based on that at the molecular scale remains a largely unsolved problem. Furthermore, experimental control over the expression of chirality at the supramolecular level is difficult to achieve^{4–7}: mixing of different enantiomers usually results in phase separation¹⁸. Here we present an experimental and theoretical description of a system in which chirality can be varied continuously and controllably ('tuned') in micrometre-scale structures. We observe the formation of twisted ribbons consisting of bilayers of gemini surfactants (two surfactant molecules covalently linked at their charged head groups). We find that the degree of twist and the pitch of the ribbons can be tuned by the introduction of opposite-handed chiral counterions in various proportions. This degree of control might be of practical value; for example, in the use of the helical structures as templates for helical crystallization of macromolecules^{8,9}.

Gemini surfactants, consisting of two identical (twin) surfactants joined by a hydrocarbon spacer of variable length, have been shown to have properties that are unusual compared to those of simple surfactants and lipids¹⁰. Cationic gemini surfactants having chiral counterions such as L-tartrate (Fig. 1) form gels in both water and some organic solvents¹¹ by creating extended networks of the multilamellar twisted ribbons reported here (Fig. 2, helix B). A feature of this system is that the chirality comes from the counterion rather than from the amphiphile itself, which allows us to both adjust the pitch and to introduce excess chirality in the form of sodium tartrate salts.

Similar structures to those that we observe are found for diacylenic lipids, bile and glutamates¹², which form long helical strips of membranes with exposed edges (Fig. 2, helix A). However, these helical ribbons are unstable: they evolve into tubules (Fig. 2) consisting of a bilayer (or multilayer) membrane of amphiphilic molecules wrapped in a cylinder which exhibits a 'barber's-pole' pattern on its surface as evidence of its chiral origin.

The twisted ribbons that we observe have several original features. Geometrically, their saddle-like curvature differs from the cylind-

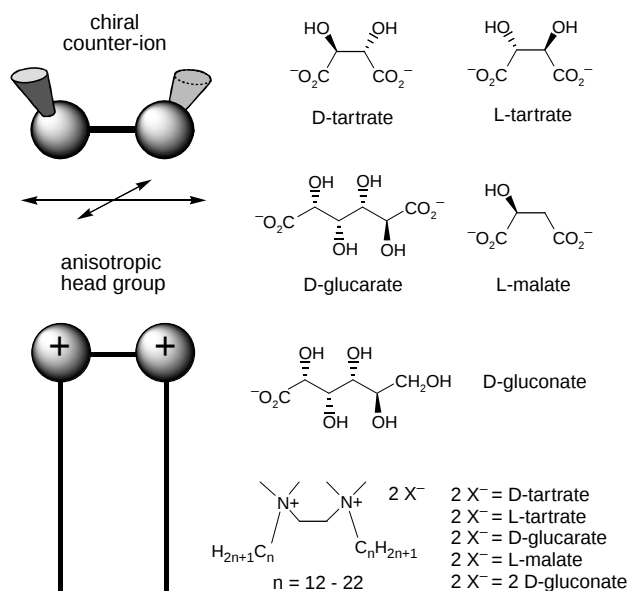


Figure 1 Cationic gemini amphiphiles having chiral counterions. Shown are the structures of ethylene-1,2-bis(dimethylalkylammonium) surfactants with various chiral anions. The nomenclature n -2- n refers to the number of carbon atoms of the first hydrocarbon chain, the ethylene spacer, and the second hydrocarbon chain, respectively. Only the tartrate derivatives form twisted ribbons, the pitch and width of which decreases for shorter hydrocarbon chains. When the counterion is L-malate, which lacks one of the hydroxy groups of tartrate, hydrogen-bonding between the counterions is weakened and only flat bilayers are observed despite the chirality of the components. When using gluconate or glucarate as counterions, extended hydrogen bonding remains possible, but the affinity of the anions for the surfactant head groups is decreased due to the larger distance between the negative charges. Again, these ions do not cause a twist in the amphiphilic bilayers.

rical curvature of helical ribbons. Also, in contrast with the tubule-forming systems, we observe no evolution towards closed chiral aggregates. The twisted ribbons may be thermodynamically stable, as is predicted for highly chiral systems by the model described below. Most importantly, we demonstrate that the chirality of the microstructures can be continuously varied by mixing of counterion enantiomers.

We find that the pure 16-2-16 L-tartrate consistently forms twisted ribbons of the same handedness, whilst the D-enantiomer forms ribbons of opposite handedness (the 16-2-16 nomenclature is defined in Fig. 1 legend). However, these enantiomers do not undergo a lateral phase separation into helices with opposite chiralities when mixed in arbitrary proportions. Instead, they mix homogeneously and form helices with a continuous variation of twist period and width (Fig. 3). An increase of the enantiomeric excess (e.e.) causes a decrease of the twist pitch from infinite (flat ribbons) to 200 nm (Table 1): e.e. is defined as the relative concentration difference $(\phi_L - \phi_D)/(\phi_L + \phi_D)$, from 0 (racemic) to 1 (pure L-tartrate). On increasing the enantiomeric excess, we also see a decrease of the mean width of the ribbons from ~400 nm to 40 nm, and their period and width seem to become more regular (Table 1). On addition of the salt sodium L-tartrate, a helix of 16-2-16 L-tartrate twists even further (Fig. 3d). The decrease of the pitch and width eventually reaches a limit (100 nm and 20 nm, respectively) at ~3 equiv. of salt added, presumably as the cationic bilayers become saturated with chiral anions (Table 1). Thus, tuning the geometrical parameters of these helices can be performed using very simple means.

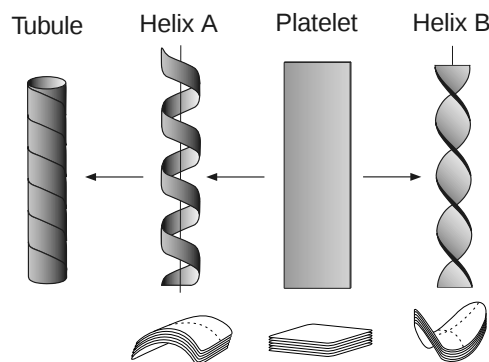


Figure 2 Schematic representation of helical and twisted ribbons. Top, platelet or flat ribbon. Helical ribbons (helix A), precursors of tubules, feature inner and outer faces. Twisted ribbons (helix B), formed by some n -2- n tartrate surfactants, have equally curved faces and a C2 symmetry axis. Bottom, the consequences of cylindrical and saddle-like curvatures in multilayered structures. In a stack of cylindrical sheets, the contact area from one layer to the next varies. This is not the case for saddle-like curvature, which is thus favoured when the layers are coordinated.

How does the shape of the twisted ribbons arise from the particular molecular structure of the amphiphiles? To answer this, we examined the properties of structural variants of 16-2-16 tartrate, and came to the following conclusions.

First, the absence of twist for the malate derivative (Fig. 1), and the properties of the gels in organic solvents¹¹, strongly support extensive hydrogen-bonding of tartrate counterions between consecutive bilayers. For simple geometric reasons, such interlayer coordination tends to favour a saddle-like curvature over that of a cylinder (see Fig. 2 legend).

Second, in water the twisted ribbons are stable for gemini surfactants with hydrocarbon chains ranging from 14 to 16 carbons. For shorter hydrocarbon chains, only micelles form for concentrations at least up to 40%. At 18 carbons, helical ribbons form instead of twisted ribbons, presumably as a result of the elevated chain melting temperature and the consequent solid phase, which will favour a cylindrical shape.

Third, the ability of the amphiphile to form such anisotropic aggregates (having ribbons with both long and short sides) appears to be directly linked to its dimeric character. For instance, the monomeric counterpart, cetyl trimethyl ammonium tartrate, fails to form any ribbon-like structure, whilst gemini surfactants having simple bromides as counterions form flat ribbons similar to those of Fig. 3a (ref. 12). Furthermore, these highly anisotropic structures strongly suggest a long-range alignment of the constituent

Table 1 Properties of ribbons formed by 16-2-6 tartrate

e.e.	T (μm)	W (μm)	T/W
0	∞	0.40 ± 0.28	
0.2	5.0 ± 1.5	0.12 ± 0.046	43 ± 5.2
0.33	2.6 ± 0.9	0.11 ± 0.047	26 ± 7.0
0.5	1.0 ± 0.5	0.06 ± 0.016	17 ± 4.7
1	0.2 ± 0.024	0.04 ± 0.006	4.8 ± 0.51
1*	0.12 ± 0.007	0.02 ± 0.002	7.0 ± 0.90
1†	0.115 ± 0.01	0.02 ± 0.002	5.6 ± 0.86
1‡	0.115 ± 0.006	0.02 ± 0.002	5.9 ± 0.63

Shown are values of enantiomeric excess (e.e., defined as $(\phi_L - \phi_D)/(\phi_L + \phi_D)$) and the mean values and standard deviations of ribbon period (T), width (W), and period/width (T/W), for different proportions of L and D enantiomers of 16-2-16 tartrate. Statistics were derived from 10–20 measurements for each sample.

* In the presence of 1 equiv. of sodium L-tartrate.

† In the presence of 3 equiv. of sodium L-tartrate.

‡ In the presence of 10 equiv. of sodium L-tartrate.

molecules, which may be correlated with a hydrogen-bonded network of tartrate ions¹¹.

In the model we propose, orientational order of the molecules within the membrane together with the chirality of the constituents provide a natural explanation for the observed twisted-ribbon shape, and can also account for the observed trends in the pitch and width of the ribbons on variation of the enantiomeric excess. In a membrane composed of gemini surfactants, an anisotropic fluid-like arrangement of head groups would be analogous to nematic order in bulk liquid crystals¹³.

The shape of the membrane is described by a curvature tensor K_{ij} , whose eigen values c_1 and c_2 are the principal curvatures along two orthogonal directions in the membrane¹⁴. Including both bending energies and the coupling of membrane shape to the assumed nematic field¹³ $Q_{ij} = S(n_i n_j - \frac{1}{2} \delta_{ij})$, the free energy per unit area of membrane is given by^{15,16}:

$$\begin{aligned} f &= \frac{1}{2} \kappa (\text{Tr} K_{ij})^2 + \bar{\kappa} \text{Det} K_{ij} + \lambda \epsilon_{ij} Q_{jk} K_{ki} + \frac{1}{2} \kappa' Q_{ij} K_{jk} K_{ki} \\ &= \frac{1}{2} \kappa (c_1 + c_2)^2 + \bar{\kappa} (c_1 c_2) + \lambda S \sin(2\theta) (c_1 - c_2) \\ &\quad + \frac{1}{2} \kappa' S \cos(2\theta) (c_1 + c_2) (c_1 - c_2) \end{aligned}$$

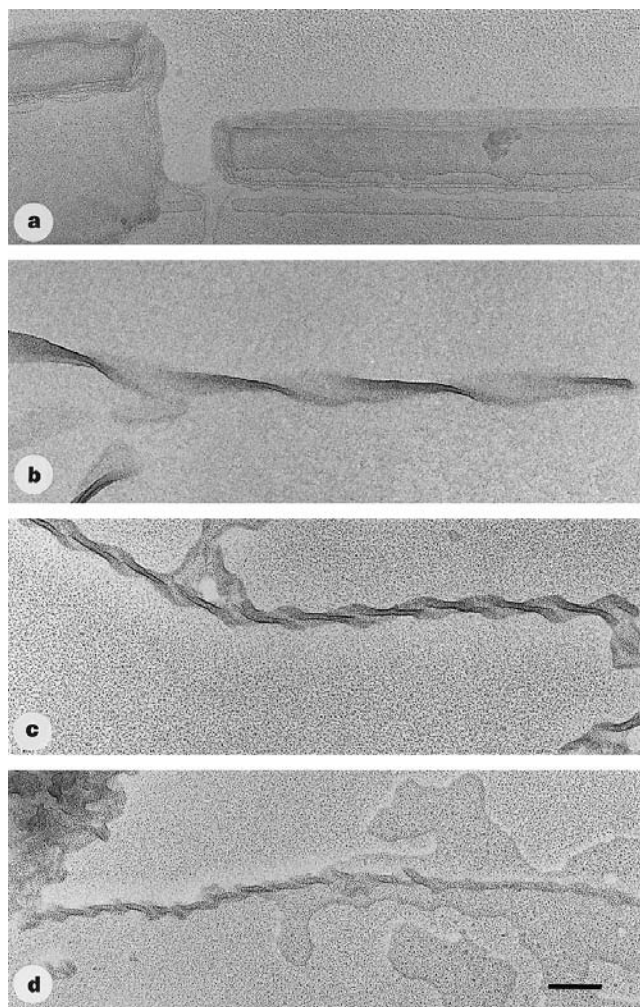


Figure 3 TEM images of the ribbons. Shown are representative twisted ribbons formed by 16-2-16 tartrate at 0.1% in water for various values of $(\phi_L - \phi_D)/(\phi_L + \phi_D)$. **a**, 0 (racemate); **b**, 0.5; **c**, 1 (pure L); **d**, 1 (pure L) in the presence of 1 equiv. of sodium L-tartrate. Scale bar, 100 nm.

Here, the coupling λ characterizes the degree of chirality of the system (which should be proportional to the enantiomeric excess), θ is the angle that the molecular orientation makes with respect to the principle curvature axes, ϵ_{ij} is the antisymmetric tensor and κ , $\bar{\kappa}$ and κ' represent the mean, gaussian and anisotropic bending stiffness. Finally, we take into account the energy cost of the exposed edges of the ribbons by introducing an excess free energy γ per unit length of edge. This model assumes that the degree of nematic order S is fixed, by, for example, detailed molecular packing. Otherwise, it represents the most general continuum model that includes the curvature free-energy terms through quadratic order. It is thus expected to be valid provided that the membrane curvature is not too large (for example, on the molecular scale).

The analysis and results of this model will be described in greater detail elsewhere. Here we summarize the most relevant results for our experiments. First, for a given width of ribbon, the optimal shape (obtained by minimizing the bulk free energy f with respect to shape) for chiral systems is that of the twisted ribbon with saddle-like curvature (specifically, $c_1 = -c_2$) and the nematic director oriented at 45° with respect to the curvature axes. For a fluid membrane, cylindrical curvature—either as a complete tubule, or as a helical ribbon precursor to a tubule—is not expected. In contrast, solid-like order would strongly oppose saddle-like curvature and favour cylindrical curvature, as is observed in tubule-forming systems¹⁷.

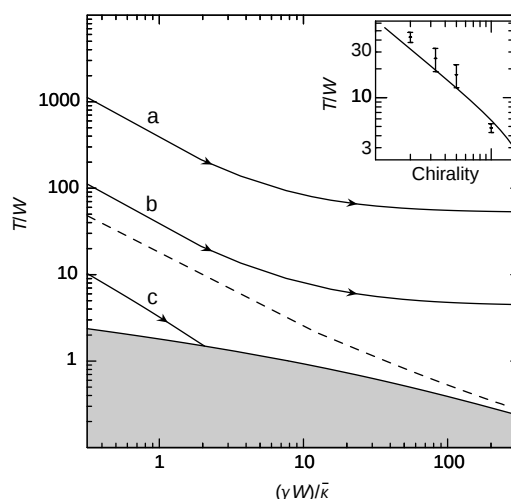


Figure 4 Predicted aspect ratio of twisted ribbons. Main figure, the calculated ratio of period to width, T/W , as a function of reduced width $\gamma W/\bar{\kappa}$. This is shown for increasing values of $\mu = \lambda S/\gamma$, which is proportional to the enantiomeric excess. Curve **a** is for $\mu = 0.02$; **b** is for $\mu = 0.2$; and **c** is for $\mu = 2$. For a given μ , only solutions lying along the corresponding line are possible. However, these are typically non-equilibrium solutions, evolving as shown by the arrows towards wider structures, which reduce the total edge energy. For narrow ribbons, the twist period $T \propto 1/\mu$ is constant. For $\mu < 1/2$ (dashed line) no equilibrium solutions are possible, although T and W become strongly correlated: the ratio T/W approaches a value near $1/\mu$, even though polydisperse distributions of both W and T are expected. For $\mu > 1/2$, the ribbons grow only until fixed, equilibrium values of T and W (shown by the filled circle) are reached. The equilibrium structures correspond to the upper boundary of the shaded region, within which no stable solutions are predicted. For all but a very narrow range of μ near the critical value of $1/2$, these are characterized by T/W of the order of 1. Inset, Measured ratio T/W (points with error bars representing standard deviations of 10–20 samples for ribbons of mixtures of 16-2-16 D and L tartrate, without any salt added. Enantiomeric excess (chirality) ranges from 0.2 to 1.0 (pure enantiomer). The curve represents the predicted steady-state values of T/W as a function of chirality $\mu = \lambda S/\gamma$. The precise relationship of the enantiomeric excess to μ involves an unknown constant of proportionality, which has been chosen to fit the data.

As depicted in Fig. 4, minimization of the total free energy per surfactant predicts equilibrium ribbons if the chirality or enantiomeric excess is large enough (specifically, if $\lambda S \geq \gamma/2$). Thus, despite the inevitable excess free energy associated with the edges, these ribbons may not grow beyond a certain preferred width W . In this case, the twist period T and ribbon width W are expected to vary inversely with the enantiomeric excess, to be well-defined (for example, their distributions are expected to be monodisperse), and the ratio T/W is expected to be of the order of 1, and nearly independent of the enantiomeric excess. Before these equilibrium structures are attained, however, it is expected that T varies inversely with the enantiomeric excess and is uncorrelated with W . Thus, the aspect ratio and monodispersity that we observe for ribbons formed by the pure 16-2-16 L-tartrate suggest that these are close to equilibrium.

The non-equilibrium structures predicted for weakly chiral systems (specifically, if $\lambda S \geq \gamma/2$) are expected to exhibit correlated values of T and W : T/W varies inversely with enantiomeric excess, although T and W are expected to have polydisperse distributions. This corresponds well to the twisted ribbons formed by mixtures of D and L enantiomers, which exhibit such correlated, though polydisperse, geometrical parameters, and large T/W ratios. The observed stability for these helices probably results from kinetic effects which limit their growth.

Thus, on the basis of molecular parameters such as chirality and anisotropy of the surfactant polar heads, this simple yet general model predicts that the observed twisted-ribbon shape is optimal, and that both the pitch and the width are expected to follow the trends that we observe on varying the chirality. Also important are the apparently qualitatively different behaviours for strongly and weakly chiral systems. The experimental observations and the theoretical model reported here may open the prospect of creating stable structures of variable pitch in amphiphilic bilayer systems. □

Received 4 January; accepted 24 March 1999.

1. Schnur, J. M. Lipid tubules: a paradigm for molecularly engineered structures. *Science* **262**, 1669–1676 (1993).
2. Fuhrop, J.-H. & Helfrich, W. Fluid and solid fibers made of lipid molecular bilayers. *Chem. Rev.* **93**, 1565–1582 (1993).
3. Rowan, A. E. & Nolte, R. J. M. Helical molecular programming. *Angew. Chem. Int. Edn* **37**, 63–68 (1998).
4. Chung, D. S., Benedek, G. B., Konikoff, F. M. & Donovan, J. M. Elastic free energy of anisotropic helical ribbons as metastable intermediates in the crystallization of cholesterol. *Proc. Natl Acad. Sci. USA* **90**, 11341–11345 (1993).
5. Spector, M. S. *et al.* Controlling the morphology of chiral lipid tubules. *Langmuir* **14**, 3493–3500 (1998).
6. Prins, L. J., Huskens, J., de Jong, E., Timmerman, P. & Reinhoudt, D. N. Complete asymmetric induction of supramolecular chirality in a hydrogen-bonded assembly. *Nature* **398**, 498–502 (1999).
7. Yashima, E., Maeda, K. & Okamoto, Y. Memory of macromolecular helicity assisted by interaction with achiral small molecules. *Nature* **399**, 449–451 (1999).
8. Ringler, P., Muller, W., Ringsdorf, H. & Brissos, A. Functionalized lipid tubules as tools for helical crystallization of proteins. *Chem. Eur. J.* **3**, 620–625 (1997).
9. Wilson-Kubalek, E. M., Brown, R. E., Celia, H. & Milligan, R. A. Lipid nanotubes as substrates for helical crystallization of macromolecules. *Proc. Natl Acad. Sci. USA* **95**, 8040–8045 (1998).
10. Zana, R. & Talmon, Y. Dependence of aggregate morphology on structure of dimeric surfactants. *Nature* **362**, 228–230 (1993).
11. Oda, R., Huc, I. & Candau, S. J. Gemini surfactants as new, low molecular weight gelators of organic solvents and water. *Angew. Chem. Int. Edn* **37**, 2689–2691 (1998).
12. Oda, R. *et al.* Elongated aggregates formed by cationic gemini surfactants. *Langmuir* **15**, 2384–2390 (1999).
13. de Gennes, P. G. & Prost, J. *The Physics of Liquid Crystals* 2nd edn (Clarendon, Oxford, 1993).
14. Helfrich, W. Elastic properties of lipid bilayers: theory and possible experiments. *Z. Naturforsch. C* **28**, 693–703 (1973).
15. Helfrich, W. & Prost, J. Intrinsic bending force in anisotropic membranes made of chiral molecules. *Phys. Rev. A* **38**, 3065–3068 (1988).
16. Selinger, J. V., MacKintosh, F. C. & Schnur, J. M. Theory of cylindrical tubules and helical ribbons of chiral lipid membranes. *Phys. Rev. E* **53**, 3804–3818 (1996).
17. Thomas, B. N., Safinya, C. R., Plano, R. J. & Clark, N. A. Lipid tubule self-assembly: length dependence on cooling rate through a first-order phase transition. *Science* **267**, 1635–1638 (1995).
18. Singh, A. *et al.* Lateral phase separation based on chirality in a polymerizable lipid and its influence on formation of tubular microstructures. *Chem. Phys. Lipids* **47**, 135–148 (1988).

Acknowledgements. This work was supported by the CNRS, INSERM, l'Hôpital Universitaire de Strasbourg, and Rhône-Poulenc. F.C.M. was supported in part by the CNRS, the Deutsche Forschungsgemeinschaft, the National Science Foundation, and the Petroleum Research Fund (administered by the ACS).

Correspondence and requests for materials should be addressed to R.O. or I.H. (e-mail: reiko.oda@iecb-polytechnique.u-bordeaux.fr).

Causes of twentieth-century temperature change near the Earth's surface

Simon F. B. Tett*, Peter A. Stott*, Myles R. Allen†, William J. Ingram* & John F. B. Mitchell*

* Hadley Centre for Climate Prediction and Research, Meteorological Office, London Road, Bracknell, Berkshire RG12 2SY, UK

† Space Science Department, Rutherford Appleton Laboratory, Chilton, OX11 0QX, UK and Department of Physics, University of Oxford

Observations of the Earth's near-surface temperature show a global-mean temperature increase of approximately 0.6 K since 1900 (ref. 1), occurring from 1910 to 1940 and from 1970 to the present. The temperature change over the past 30–50 years is unlikely to be entirely due to internal climate variability^{2–4} and has been attributed to changes in the concentrations of greenhouse gases and sulphate aerosols⁵ due to human activity. Attribution of the warming early in the century has proved more elusive. Here we present a quantification of the possible contributions throughout the century from the four components most likely to be responsible for the large-scale temperature changes, of which two vary naturally (solar irradiance and stratospheric volcanic aerosols) and two have changed decisively due to anthropogenic influence (greenhouse gases and sulphate aerosols). The patterns of time/space changes in near-surface temperature due to the separate forcing components are simulated with a coupled atmosphere–ocean general circulation model, and a linear combination of these is fitted to observations. Thus our analysis is insensitive to errors in the simulated amplitude of these responses. We find that solar forcing may have contributed to the temperature changes early in the century, but anthropogenic causes combined with natural variability would also present a possible explanation. For the warming from 1946 to 1996 regardless of any possible amplification of solar or volcanic influence, we exclude purely natural forcing, and attribute it largely to the anthropogenic components.

The coupled model we use is HadCM2 (refs 6, 7), which has a horizontal resolution of 2.5° in latitude by 3.75° in longitude, 19 atmospheric and 20 oceanic levels, and a flux correction. Its climate sensitivity to a doubling of the atmospheric concentration of CO₂ is estimated to be 3.3 K (C. A. Senior, personal communication).

The main radiative forcings of climate since 1850 are likely to be anthropogenic changes in well-mixed greenhouse gases and tropospheric aerosols (mainly sulphate), and natural changes in solar irradiance and in stratospheric aerosol due to volcanic activity⁸. We compare observations¹ of 10-year mean near-surface temperature changes over five 50-year periods (1906–56, 1916–66, ..., 1946–96) with simulations of HadCM2 forced by the following factors (see also section 1 of Supplementary Information):

'G'. Changes in well-mixed greenhouse gases from 1860 to 1996^{8,9,10} expressed as equivalent CO₂.

'GS'. As G but also with changes in surface albedo¹¹ representing the effects of anthropogenic sulphate aerosols from 1860 to 1996^{6,9,10} derived from an atmospheric chemistry model¹². We assume that this albedo represents both the direct and indirect^{11,13,14} effects of sulphate aerosols. When we consider both G and GS we define a further signal S, the pure sulphate signal, as GS – G (see section 9 of Supplementary Information).

'Vol'. Changes in stratospheric volcanic aerosols¹⁵ from 1850 to 1996.

'Sol'. Changes in total solar irradiance from 1700 to 1996 based on proxy data¹⁶ for 1700–1991 and extended to 1996 using satellite observations¹⁷.

The climate response to each of the above factors was computed from the ensemble average of four simulations started from different initial conditions. Internal climate variability is estimated from a 1,700-year control simulation (Control). 50-year segments (of 10-year mean near-surface temperatures) from the responses of the four ensembles (signals), Control and observations were processed identically to allow for the effect of changing observational coverage, to filter out variability on scales less than 5,000 km (ref. 18) and to remove the 50-year time-mean (see section 2 of Supplementary Information).

We assume that the observations (y) may be represented as the sum of simulated responses or signals (x_i) and internal climatic variability (u), which we assume to be normally distributed:

$$y = \sum_i \beta_i x_i + u \quad (1)$$

This assumption of linearity has been shown to be a good approximation in at least one general circulation model^{19,20}. The amplitude, β_i , represents the amount by which we have to scale the i th signal to give the best fit to the observations. We use an “optimal fingerprinting” algorithm^{21–24}, a form of multivariate regression, to estimate these amplitudes and uncertainty ranges (see section 3 of Supplementary Information). Here we use 5–95% uncertainty ranges and if the uncertainty range for β_i includes unity, then the amplitude of the simulated signal (x_i) could be correct. The first millennium of Control is used to define a weighting function which minimizes the influence of patterns of high internal variability on the estimated signal amplitudes (giving “optimal fingerprints”). Observations and all simulated data are then further filtered by projection on to the leading ten modes of spatio-temporal variability from this part of Control (see section 7 of Supplementary Information). The latter seven centuries, which are statistically independent, are used to estimate the uncertainty ranges.

A minimum requirement for a signal or signal-combination to explain temperature changes over a 50-year period is that the amplitude of the estimated residual ($\Sigma \beta_i x_i - y$) should be consistent, at the 5% level, with internal climate variability²⁴ estimated from Control. For a signal-combination to explain change over the century we require that the estimated residual be consistent in all five 50-year periods. We test this using a F -test (see sections 4 and 6 of Supplementary Information) which rules out simulated internal variability as an explanation of temperature changes during 1906–56, 1916–66 and 1946–96 (Table 1), despite being a relatively weak test in that it makes no use of the shape of the simulated signals. Furthermore, no combination of our simulated natural signals alone can explain the warming since 1946—the period when the

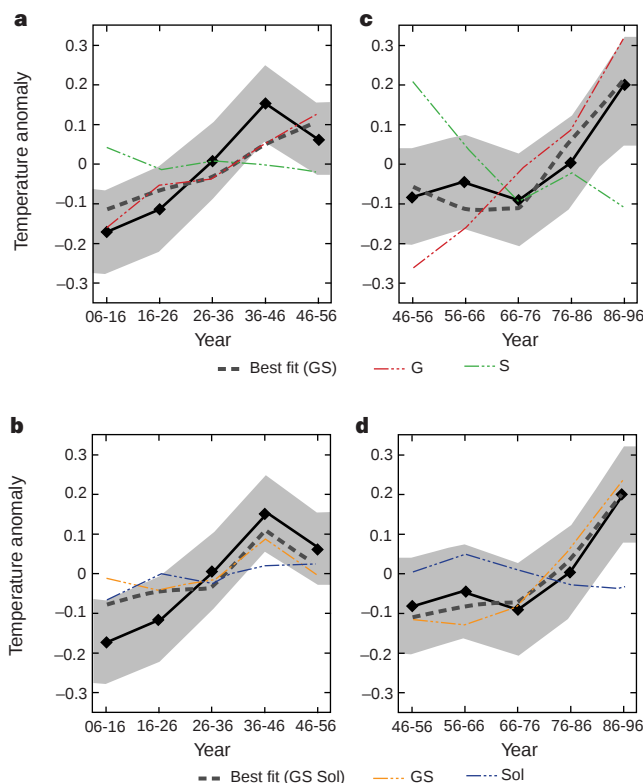


Figure 1 Best-estimate contributions to global-mean temperature change in the twentieth century. Shown are reconstructions of temperature variations for 1906–56 (**a** and **b**) and 1946–96 (**c** and **d**) for G&S (**a** and **c**) and GS&Sol (**b** and **d**). Observed (thick black lines), best fit (dark grey dashed lines), and the uncertainty range due to internal variability (grey shading) are shown in all plots. Panels **a** and **c** show contributions from G (dashed red) and S (dashed green); panels **b** and **d** show the contributions from GS (dashed orange) and Sol (dashed blue). All time series were reconstructed from data in which the 50-year mean had first been removed.

observational coverage is best. However, every case that includes an anthropogenic signal passes the consistency test. These results hold even if the global-mean model sensitivity or the amplitude of the forcing is wrong as we place no restriction on the amplitudes of the simulated signals.

A risk in multivariate linear regression is signal degeneracy, which means that signals resemble linear combinations of one another. Applying tests for degeneracy (see pages 243–248 of ref. 25 and section 8 of Supplementary Information) to our signals and observations, we find that no more than two signal-amplitudes can be estimated simultaneously. While three-signal and four-signal combinations are possible explanations, results from them may be misleading. Thus we only discuss combinations of at most two signals.

A signal is formally detected in a signal-combination and period if that combination is consistent with observations and the uncertainty range for the signal is entirely positive (see sections 5 and 6 of Supplementary Information). This means that we have rejected the null hypothesis that the signal amplitude is zero (or that the simulated signal is of the wrong sign). In other words, it is highly likely that the signal is present in the observations.

All one-or two-signal combinations that include anthropogenic signals are, on the basis of the consistency test, possible explanations of twentieth-century temperature change, but we wish to focus on the ‘best’ explanation. Ideally, this explanation would include all detectable signals and no others, provided that this combination is consistent with the observations. We find that three signals (G, S and Sol) are detected during the century (Table 1). The composite signal

Table 1 Consistency tests

	<i>F</i> -test*				
	1906–56	1916–66	1926–76	1936–86	1946–96
Int. var.†	0.01	0.04	0.69	0.76	0.01
G	0.18 ^G	0.11	0.75	0.80	0.10 ^G
GS	0.06 ^{GS}	0.23	0.76	0.85	0.30 ^{GS}
Sol†	0.11 ^{Sol}	0.06	0.75	0.87	0.03
Vol†	0.03	0.08	0.74	0.80	0.04†
G&S	0.21 ^G	0.17 ^G	0.68	0.83	0.31 ^{G,S}
G&Sol	0.23 ^G	0.17 ^{G,Sol}	0.68	0.81	0.07 ^G
G&Vol	0.24 ^G	0.09	0.67	0.73	0.08 ^G
GS&Sol	0.13 ^{Sol}	0.19 ^{GS}	0.68	0.82	0.25 ^{GS}
GS&Vol	0.05 ^{GS}	0.22 ^{GS}	0.68	0.80	0.25 ^{GS}
Sol&Vol†	0.08 ^{Sol}	0.09 [‡]	0.67	0.82	0.04‡

* These columns show the probability (P) that the best-fit signal combination is consistent with observations using an F -test with 21 degrees of freedom. Values in bold are inconsistent with the observations. When a signal is detected, its name is shown as a superscript next to the P -value. Actual P -values are shown, although 0.03 is the lowest value that can be robustly estimated given the available length of Control.

† Signal combination is an inadequate explanation of twentieth-century temperature change as the F -test fails at least once.

‡ Volcanic amplitude in this signal-combination is significantly negative and thus unphysical. No other signal ever has a significant negative amplitude.

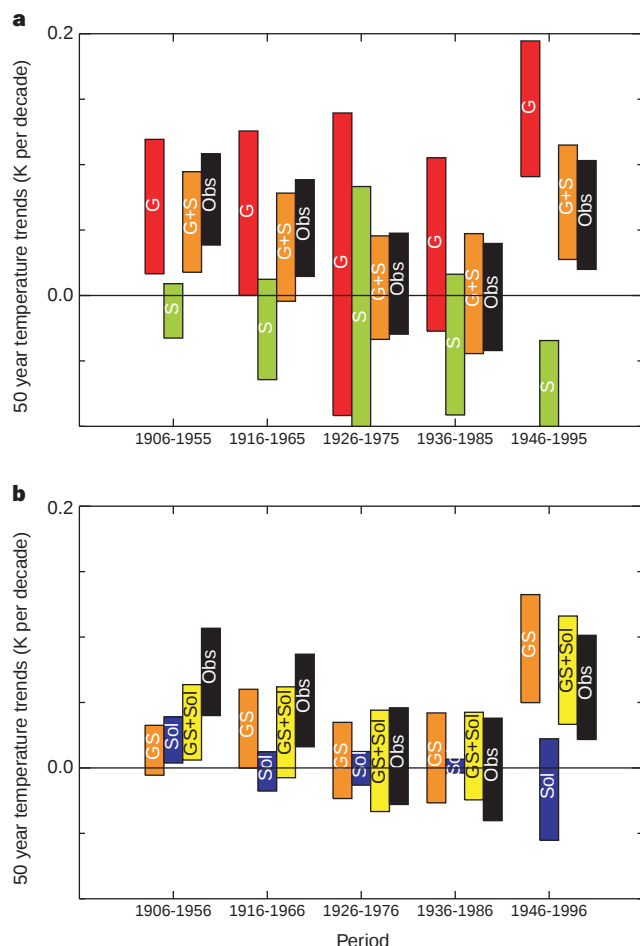


Figure 2 50-year temperature trends, and uncertainties, due to different causes. Black bars, observed 50-year trends (K per decade) in global-mean near-surface temperature with 5–95% range estimated from Control. **a**, Estimated contribution (and uncertainty range) to observed trends from G (red), S (green) and the sum (orange). **b**, Estimated contribution and uncertainty range from GS (orange), Sol (blue) and total (yellow). The signal name is centred at the estimated value and the bars cover the uncertainty range.

GS (in which the relative amplitudes of the greenhouse and sulphate signals are assumed to be as in the GS experiment, giving a single pattern of response to anthropogenic forcing) can also be detected. As signal degeneracy may make results from the three-signal combination, G&S&Sol, misleading, this leaves three two-signal combinations: G&S, GS&Sol and G&Sol. Before the most recent period, none of these three two-signal combinations are an obviously better fit to the observations than any other (Table 1). But in 1946–96, combinations including the influence of sulphate aerosols (S) fit the observations better than other explanations (Table 1), which all fail to pass the *F*-test at the 10% level, suggesting that some sulphate influence is required to account for recent changes. Thus we consider the two combinations containing S (G&S and GS&Sol), but we explore the sensitivity of the solar detection during 1906–56 to the size of the sulphate signal included in GS.

In G&S (unlike GS), the observations are allowed to determine the relative amplitude of the sulphate and greenhouse-gas signals. This allows for error in the prescribed sulphate forcing or modelled response which varies with each period. With these two signals we detect S in 1946–96 and G in 1906–56, 1916–66 and 1946–96 (Table 1). We find that during 1906–56 the amplitude of the sulphate aerosol signal relative to the greenhouse-gas signal is significantly less than that prescribed in GS.

Some other factor may be required to explain this discrepancy with GS rather than errors in the simulated anthropogenic signals. If the relative amplitude of the anthropogenic signals is as prescribed in the GS experiment, then Sol is detected: during 1906–56, some solar influence is required to explain the temperature changes (Table 1). If we reduce the amplitude of the sulphate signal in GS by 33% or more (making GS more like G), we detect the anthropogenic signal rather than the solar signal during 1906–56. GS is detected in 1946–96 with or without this change. The reduction of the relative amplitude of sulphate aerosols to greenhouse gases is within the possible uncertainty range⁸; thus our detection of a solar signal should be treated with some caution.

We reconstruct global-mean temperature changes by multiplying the simulated signals, x_i by the best-fit amplitudes (β_i) for 1906–56 and 1946–96. During 1906–56, in G&S, greenhouse gases warm, counterbalanced by a weaker aerosol effect than was imposed in GS (Fig. 1a). The warming peak around 1940 is accounted for by a combination of internal variability and a steadily increasing temperature due to anthropogenic forcings. In GS&Sol the observed early-century warming is explained by a combination of solar irradiance changes, internal variability and an increase in greenhouse gases (partly balanced by the radiative effect of sulphate aerosols), with solar activity largely responsible for the warming peak around 1940 (Fig. 1b). In both cases, from 1946 sulphate aerosols balance the effect of greenhouse gases giving little warming until the mid-1970s when the warming due to increasing greenhouse gases dominates (Fig. 1c, d).

In GS&Sol, solar influence is detectable during the 1906–56 period (Fig. 2), but contributes little to the linear trend in global-mean temperature (a best estimate of ~ 0.025 K per decade). Conversely, G&S shows that greenhouse gases could have contributed substantially to the warming during both the 1906–56 and 1916–66 periods (Fig. 2). In all other periods, the best-estimate solar contribution is negligible. For both cases the best estimate of the total anthropogenic warming is ~ 0.075 K per decade during 1946–96, within an approximate uncertainty range of 0.03–0.11 K per decade.

To test the robustness of our results, we carried out six sensitivity studies in which we changed our analysis procedure (see section 10 of Supplementary Information). Detection of anthropogenic signals during 1906–56 (in G&S) and 1946–96 (in both cases) was unaltered by those changes. Detection of the 1946–96 sulphate signal is significant only at the 90% level if optimization is not used. Detection of the 1906–56 solar signal is significant at the 90% level only if greater weight is given to smaller spatial scales, but remains significant at the 95% level in all other studies. We also used a different solar irradiance reconstruction²⁶ from 1750, to produce an alternative Sol signal, and we reduced stratospheric ozone from 1974⁴ to modify the GS signal. Our principal results are robust to both these changes. Furthermore, if we inflate the variance of Control by a factor of 4 we still detect an anthropogenic signal in the 1946–96 period in both of the two-signal cases considered.

Our simulations did not explicitly represent the effects of sulphate aerosols on cloud albedo¹³ or lifetime¹⁴ (though on the scales considered these are likely to be represented by the albedo changes imposed on the model), nor did we consider the climate effects of other aerosols, changes in the spectral distribution of the solar irradiance and possible associated changes in ozone^{27,28}. Although our results are unaffected by error in the amplitude of the forcing, they could be sensitive to error in the patterns of response or other errors in the forcing, though our conclusions were unaltered by use of an alternative solar irradiance reconstruction. The test that we use to evaluate consistency between the observations and simulations could be misled by mutually compensating errors in model estimates of natural variability and the response of the model to external factors. We do not consider possible observational errors (likely to be small relative to internal climate variability) nor has our

uncertainty analysis included any uncertainty due to the finite length of data used to derive the optimisation. For consistency with earlier work^{21–24}, we have used standard estimates of pattern amplitudes based on linear regression which are biased towards zero²⁵ when, as here, there is uncertainty in the signals.

Bearing in mind these caveats, we interpret our results as showing the following: first, the temperature changes over the twentieth century cannot be explained by any combination of natural internal variability and the response to natural forcings alone. Second, the recent warming, ~ 0.25 K, can be explained by the response of the climate to anthropogenic changes in greenhouse-gas concentrations partly offset by cooling due to anthropogenic sulphate aerosols, resulting in little net temperature change from 1946 to the mid-1970s. Last, the warming early this century can also be explained by anthropogenic causes and internal variability. However, solar irradiance changes could have made a significant contribution, ~ 0.125 K, if we assume little error in the relative amplitude of the forcing of sulphates and greenhouse gases prescribed in our model. □

Received 5 October 1998; accepted 25 March 1999.

- Parker, D. E., Jones, P. D., Folland, C. K. & Bevan, A. Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophys. Res.* **99**, 14373–14399 (1994).
- Stouffer, R. J., Manabe, S. & Vinnikov, K. Y. Model assessment of the role of natural variability in recent global warming. *Nature* **367**, 634–636 (1994).
- Santer, B. D. *et al.* A search for human influences on the thermal structure of the atmosphere. *Nature* **382**, 39–45 (1996).
- Tett, S. F. B., Mitchell, J. F. B., Parker, D. E. & Allen, M. R. Human influence on the atmospheric vertical temperature structure: Detection and observations. *Science* **247**, 1170–1173 (1996).
- Hegerl, G. C. *et al.* Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Clim. Dyn.* **13**, 613–634 (1997).
- Johns, T. C. *et al.* The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation. *Clim. Dyn.* **13**, 103–134 (1997).
- Tett, S. F. B., Johns, T. C. & Mitchell, J. F. B. Global and regional variability in a coupled AOGCM. *Clim. Dyn.* **13**, 303–323 (1997).
- Shine, K. P., Fouquart, Y., Ramaswamy, V., Solomon, S. & Srinivasan, J. in *Climate Change 1995: the Science of Climate Change* (eds Houghton, J. T. *et al.*) 108–118 (Cambridge Univ. Press, 1996).
- Mitchell, J. F. B., Johns, T. C., Gregory, J. M. & Tett, S. F. B. Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature* **376**, 501–504 (1995).
- Mitchell, J. F. B. & Johns, T. C. On modification of global warming by sulfate aerosols. *J. Clim.* **10**, 245–267 (1997).
- Mitchell, J. F. B., Davis, R. A., Ingram, W. J. & Senior, C. A. On surface temperature, greenhouse gases, and aerosols: Models and observations. *J. Clim.* **10**, 2364–2386 (1995).
- Langner, J. & Rodhe, H. A global three-dimensional model of the tropospheric sulfur cycle. *J. Atmos. Chem.* **13**, 225–263 (1991).
- Twomey, S. A. Pollution and the planetary albedo. *Atmos. Environ.* **8**, 1251–1256 (1974).
- Albrecht, B. A. Aerosols, cloud microphysics and fractional cloudiness. *Science* **245**, 1227–1230 (1989).
- Sato, M., Hansen, J. E., McCormick, M. P. & Pollack, J. B. Stratospheric aerosol optical depths (1850–1990). *J. Geophys. Res.* **98**, 22987–22994 (1993).
- Hoyt, D. V. & Schatten, K. H. A discussion of plausible solar irradiance variations, 1700–1992. *J. Geophys. Res.* **98**, 18895–18906 (1993).
- Willson, R. C. Total solar irradiance trend during solar cycles 21 and 22. *Science* **277**, 1963–1965 (1997).
- Stott, P. A. & Tett, S. F. B. Scale-dependent detection of climate change. *J. Clim.* **11**, 3282–3294 (1998).
- Haywood, J., Stouffer, R., Wetherald, R., Manabe, S. & Ramaswamy, V. Transient response of a coupled model to estimated changes in greenhouse gas and sulfate concentrations. *Geophys. Res. Lett.* **24**, 1335–1338 (1997).
- Ramaswamy, V. & Chen, C.-T. Linear additivity of climate response for combined albedo and greenhouse perturbations. *Geophys. Res. Lett.* **24**, 567–570 (1997).
- Hasselmann, K. Optimal fingerprints for the detection of time-dependent climate change. *J. Clim.* **6**, 1957–1971 (1993).
- Hasselmann, K. Multi-pattern fingerprint method for detection and attribution of climate change. *Clim. Dyn.* **13**, 601–611 (1997).
- North, G. R., Kim, K.-K., Shen, S. S. P. & Hardin, J. W. Detection of forced climate signals. Part I: filter theory. *J. Clim.* **8**, 401–408 (1995).
- Allen, M. R. & Tett, S. F. B. Checking for model consistency in optimal fingerprinting. *Clim. Dyn.* (in the press).
- Mardia, K. V., Kent, J. T. & Bibby, J. M. *Multivariate Analysis* (Academic, London, 1979).
- Lean, J., Beer, J. & Bradley, R. Reconstruction of solar irradiance since 1610: implications for climate change. *Geophys. Res. Lett.* **22**, 3195–3198 (1995).
- Haigh, J. D. Impact of solar variability on climate. *Science* **272**, 981–984 (1996).
- Haigh, J. D. The role of stratospheric ozone in modulating the solar radiative forcing of climate. *Nature* **370**, 544–546 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. S.F.B.T., P.A.S. and computer time were funded by the Department of the Environment, Transport and the Regions. W.J.L. and J.F.B.M. were supported by the UK Public Meteorological Service Research and Development programme. M.R.A. was supported by a research fellowship from the UK Natural Environment Research Council. Supplementary support was provided by the European Commission.

Correspondence and requests for materials should be addressed to S.F.B.T. (3-mail: sfbtett@meto.gov.uk).

Changing spatial structure of the thermohaline circulation in response to atmospheric CO₂ forcing in a climate model

Richard A. Wood, Ann B. Keen, John F. B. Mitchell & Jonathan M. Gregory

Hadley Centre for Climate Prediction and Research, Meteorological Office, London Road, Bracknell RG12 2SY, UK

The heat transported northwards by the North Atlantic thermohaline circulation warms the climate of western Europe^{1–3}. Previous model studies^{4–6} have suggested that the circulation is sensitive to increases in atmospheric greenhouse-gas concentrations, but such models have been criticised for the use of unphysical 'flux adjustments'^{7–9} (artificial corrections that keep the model from drifting to unrealistic states), and for their inability to simulate deep-water formation both north and south of the Greenland–Iceland–Scotland ridge, as seen in observations^{10,11}. Here we present simulations of today's thermohaline circulation using a coupled ocean–atmosphere general circulation model without flux adjustments. These simulations compare well with the observed thermohaline circulation, including the formation of deep water on each side of the Greenland–Iceland–Scotland ridge. The model responds to forcing with increasing atmospheric greenhouse-gas concentrations by a collapse of the circulation and convection in the Labrador Sea, while the deep-water formation north of the ridge remains stable. These changes are similar in two simulations with different rates of increase of CO₂ concentrations. The effects of increasing atmospheric greenhouse-gas concentrations that we simulate are potentially observable, suggesting that it is possible to set up an oceanic monitoring system for the detection of anthropogenic influence on ocean circulation.

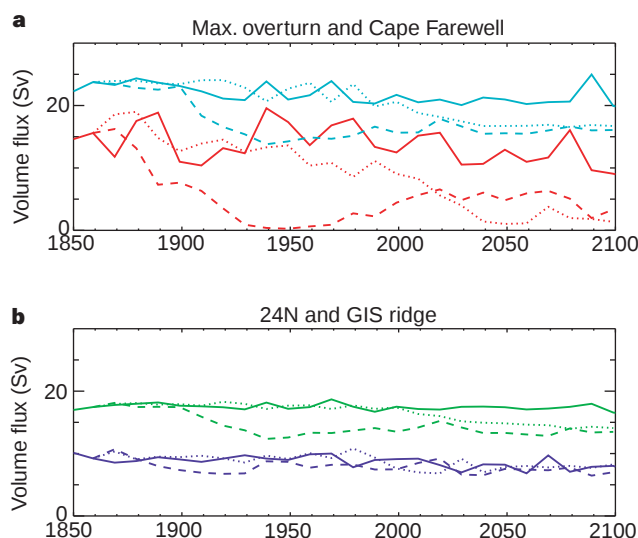


Figure 1 Time series of various components of the thermohaline circulation. **a**, Maximum meridional overturning at all latitudes in the North Atlantic (light blue), and volume flux of water denser than $\sigma_\theta = 27.5$ across Cape Farewell (red); **b**, maximum meridional overturning at 24°N (green), and volume flux of water denser than $\sigma_\theta = 27.5$ across the GIS ridge (dark blue). These time series are shown for the control (solid lines), 2PC (dashed lines) and GHG (dotted lines) runs. The GIS ridge, Cape Farewell and 24°N sections are shown in Fig. 4a. The initial 100 years of spinup of the control run are not shown.

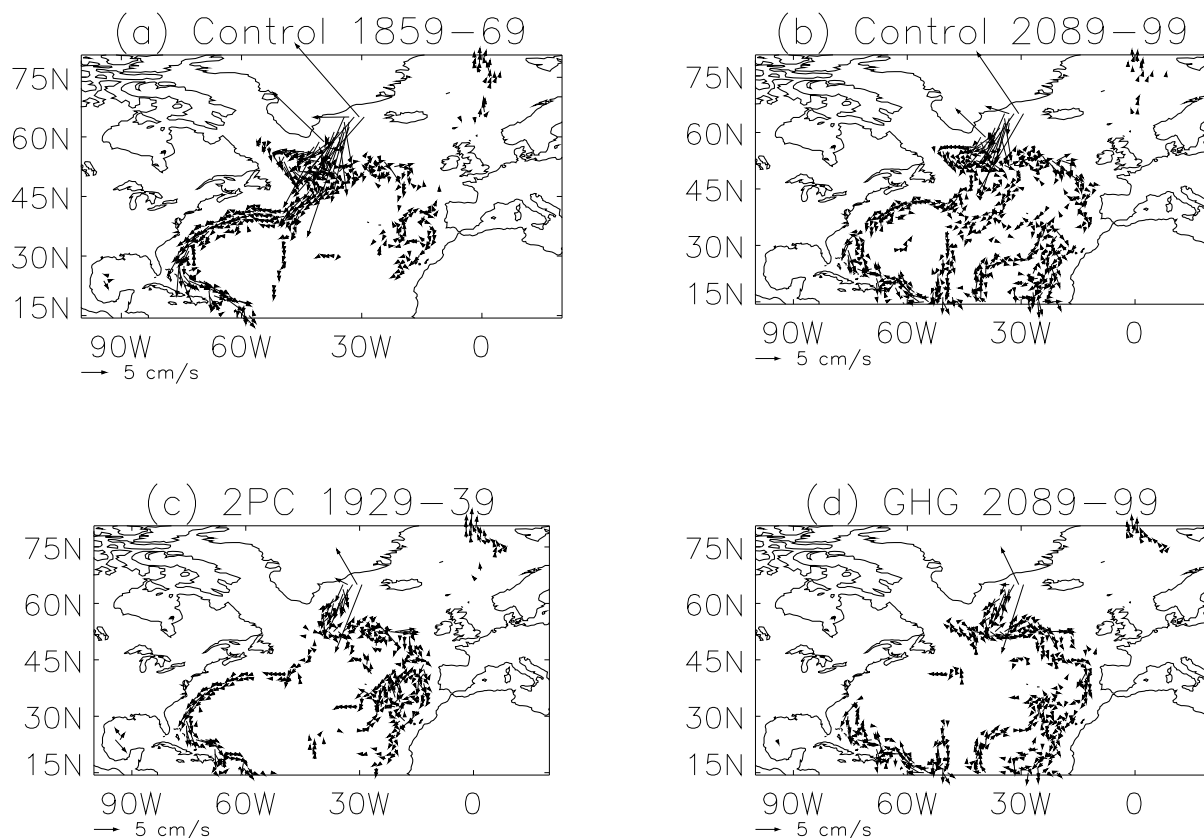


Figure 2 Decadal mean velocity in the North Atlantic at 2,731 m depth. **a**, Control run 1859–69; **b**, control run 2089–99; **c**, 2PC 1929–39; **d**, GHG 2089–99. For clarity,

only velocities greater than 0.7 cm s^{-1} are shown. A scale vector is shown at the bottom left of each panel.

The model that we use, HadCM3, is a coupled atmosphere–ocean–sea ice general circulation model developed from an earlier version used extensively for climate studies¹². Specific changes include increased ocean horizontal resolution (from 3.75×2.5 degrees longitude $\times$ latitude to 1.25×1.25 degrees), and inclusion of a modified convection scheme in regions of dense overflows¹³ and the Gent–McWilliams¹⁴ adiabatic diffusion scheme. The increased resolution allows us to represent the channels in the Greenland–Iceland–Scotland (GIS) ridge through which the dense overflow water from the Greenland, Iceland and Norwegian (GIN) seas passes into the North Atlantic. These overflows are known to influence the structure of the whole North Atlantic circulation¹⁵. Atmospheric model improvements include a new radiation scheme, and revisions to the cloud, convection, boundary layer and land surface schemes. See ref. 16 for model details.

To investigate the sensitivity of the thermohaline circulation to changes in atmospheric concentrations of greenhouse gases we perform three integrations of the model. The control simulation begins from the present-day ocean state¹⁷ and is run for 340 years, with greenhouse-gas concentrations fixed at pre-industrial values. The nominal start date of the run is 1759, and results are presented below for years 1859–2100 (that is, following a 100-year spinup period). Runs 2PC and GHG begin from the control state at 1859 and are forced with different scenarios of increasing atmospheric CO_2 . In 2PC, CO_2 is increased by 2% per year until 1929 (when it reaches four times the control value), and is held constant thereafter. This represents an unrealistically fast CO_2 increase, to highlight the sensitivity of the circulation. Run GHG follows historical concentrations of all anthropogenic greenhouse gases from 1859 to the present, followed by the IS92a ‘business as usual’ concentration scenario to 2100¹⁸. Calendar dates are only significant for run GHG.

We first describe the control simulation. The model’s large-scale

ocean heat transports agree with observational estimates and with those implied by the surface heat fluxes produced by the atmospheric component alone¹⁶, allowing a stable, realistic surface climate without use of flux adjustments^{16,19,20}. The maximum overturning in the North Atlantic has a mean value of 21.6 Sv over the period 1859–2100, with little secular trend (Fig. 1a). Although this quantity is frequently used by modellers as a measure of the strength of the thermohaline circulation, it cannot in practice be estimated from observations. We therefore diagnose the formation and transport of North Atlantic Deep Water at three important locations where there are relatively robust observations: the GIS ridge (GIN seas overflow), south of Cape Farewell (the southern tip of Greenland), and the transatlantic section at 24°N . For the GIS ridge the mean flow is 8.5 Sv, compared with the observational estimate¹¹ of 5.6 Sv (Fig. 1b). The model North Atlantic is somewhat too salty, and hence the overflow density of $\sigma_\theta = 28.3$ is higher than the observed value¹¹ of 28.0. At Cape Farewell, the transport shows interdecadal variability with a magnitude and timescale which is consistent with a recent reconstruction based on historical hydrographic data²¹ (Fig. 1a). The model’s mean transport is 13.0 Sv, consistent with the only direct current-meter measurements at this section²². At 24°N the model flow is 17.4 Sv, compared with the observational estimate² of 19.3 Sv (Fig. 1b). Thus the model gives a reasonable quantitative simulation of the flow of North Atlantic Deep Water, and in particular the partition of the flow between sources to the north and south of the GIS ridge. None of the section transports shows marked secular drift over the period 1859–2100, although there is some drift in the vertical structure of the flow at 24°N due to gradual intrusion of Antarctic Bottom Water¹⁶.

Next we discuss results from run 2PC. The maximum overturning (occurring at about 30°N) falls from around 23 Sv to 14 Sv as CO_2 is increased, then recovers slightly after 1929 (Fig. 1a). This is

qualitatively similar to some earlier studies^{5,6}, but the circulation does not collapse completely as it did in an earlier $4 \times \text{CO}_2$ simulation using the GFDL model⁴. This difference may be linked to a weaker increase of freshwater input to the North Atlantic in HadCM3 than in the GFDL model. The 'hydrological sensitivity'²³ k of HadCM3 is 0.023 Sv K^{-1} , compared with 0.03 Sv K^{-1} for the GFDL model²³. For HadCM3, k includes a contribution of 0.003 Sv K^{-1} from increased melting of the Greenland ice sheet. This agrees well with estimates based on more sophisticated ice-sheet models²⁴. Unlike previous studies, there is no possibility here that the circulation is artificially stabilized or destabilized through a large proportion of the surface heat and freshwater flux being contained in (fixed) flux adjustment terms, which cannot vary as the forcing and circulation change⁹. For this reason, we believe that we can have more confidence in the mechanisms of thermohaline circulation change produced by our model.

The changes in maximum overturning are reflected in the overturning at 24°N which drops by 20–25% by the time of CO_2 quadrupling (Fig. 1b). The volume flux across the GIS ridge is relatively stable throughout the run (Fig. 1b), although the outflow water becomes warmer and lighter as the climate warms. The main changes in the deep circulation take place south of the GIS ridge. The cyclonic circulation in the Labrador Sea (Fig. 2a, b) collapses completely by the time of CO_2 quadrupling (Fig. 2c). This is reflected in a collapse of the flow across the Cape Farewell section (Fig. 1a). The deep Labrador Sea circulation is driven by the density contrast between the dense overflow water entering the Labrador

basin at Cape Farewell and the lighter water in the interior Labrador basin. As a result of surface warming in the GIN seas, the GIS overflow water becomes lighter through the run (mean density $\sigma_\theta = 28.3$ at the start of the run to $\sigma_\theta = 28.1$ in 1929), and this is reflected downstream in the density of the deep boundary current northeast of Cape Farewell (Fig. 3a, c, segment CD). The interior Labrador water (Fig. 3a, c, segment BC) is less strongly ventilated, so its density changes less; by 1929 the deep boundary current water is no denser than the interior water, so there is no longer a pressure gradient to drive the deep water around the Labrador Sea.

Further northwest in the Labrador Sea is a site of convection, shown by the doming of density surfaces in Fig. 3a, segment AB. Labrador Sea water formed here contributes to the southward flowing North Atlantic Deep Water at 24°N . In the control run, the convection varies on interdecadal timescales, as observed²⁵ (compare Figs 3a and b), but by 1929 in 2PC (Fig. 3c) the convection has ceased as a result of surface freshening and warming.

To summarize, the amount of overflow of deep water across the GIS ridge appears remarkably insensitive to the CO_2 increase, but because the overflow water becomes lighter than the interior deep Labrador Sea its circulation around the Labrador Sea collapses. Convection in the Labrador Sea also stops. These changes are reflected further downstream at 24°N , where the NADW flow is reduced by $\sim 20\%$.

In GHG, the imposed radiative forcing at 2089 is approximately equivalent to CO_2 quadrupling, and the flow changes are qualitatively and quantitatively similar to those in 2PC in 1929 (Figs 1, 2d, 3d). Thus the circulation changes appear robust, and the final state appears to be not strongly dependent on the time profile of greenhouse-gas increase in this case, at least up to 2100^{26} . In GHG, the Labrador Sea convection collapse occurs around 1999–2029 (we note that GHG does not include all forcings, for example, the mitigating effect of sulphate aerosols which would delay the response).

By the time of CO_2 quadrupling, the sea surface temperature has increased by several degrees over most of the North Atlantic (Fig. 4), and the climate of western Europe has warmed substantially. However the ocean warming is a minimum, with even some regions of cooling (in the northeast Atlantic and Greenland Sea), suggesting that the changes in thermohaline circulation are important in modifying the surface response to the greenhouse-gas increase.

A number of factors should be borne in mind concerning our results. First, there is a slight drift in the density field through the control run (Fig. 3a, b); but this drift is smaller than the density changes in the overflow water (Fig. 3c, d). Second, the model's GIS overflows take place through grid-scale channels. Most of the model overflow passes through the Denmark Strait between Greenland and Iceland, whereas a significant proportion of this flow is observed to the east of Iceland¹¹. In reality, the channels are narrower than the model grid scale, but previous model studies suggest that the sill depths, and not their widths, may control the flow²⁷. It is clearly important to improve understanding of the physics and modelling of overflow processes, given their fundamental influence on the thermohaline circulation^{15,28}. Third, because the deep-water formation is localized in the model (as in reality), the changes in thermohaline circulation may depend on relatively small-scale features of the atmospheric response, which vary between general circulation models²⁴. Last, the sensitivity of the hydrological cycle to climate change is an important factor in the response of the thermohaline circulation, and remains subject to both observational and modelling uncertainty²³.

Despite the above uncertainties, we believe that the HadCM3 model has reached a level of realism where it can give valuable information on possible future changes in thermohaline circulation. This not only leads to more confident predictions of climate change, but can also inform the design of monitoring systems for oceanic climate change. Our results suggest a clear spatial pattern

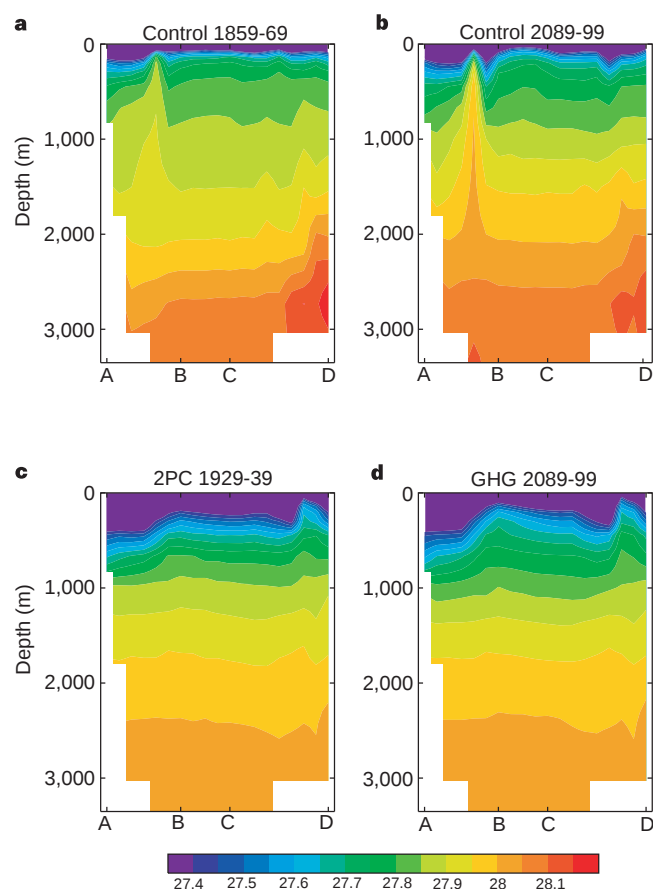


Figure 3 Potential density σ_θ (kg m^{-3}) on section ABCD. The section is shown in Fig. 4a. The segment DC follows approximately the core of the dense overflow water from Denmark Strait. The doming between A and B is evidence of convection in the Labrador Sea. There is some drift in the control run on this section (compare **a** and **b**) but the stronger doming in **b** is partly due to interdecadal variability in the control run.

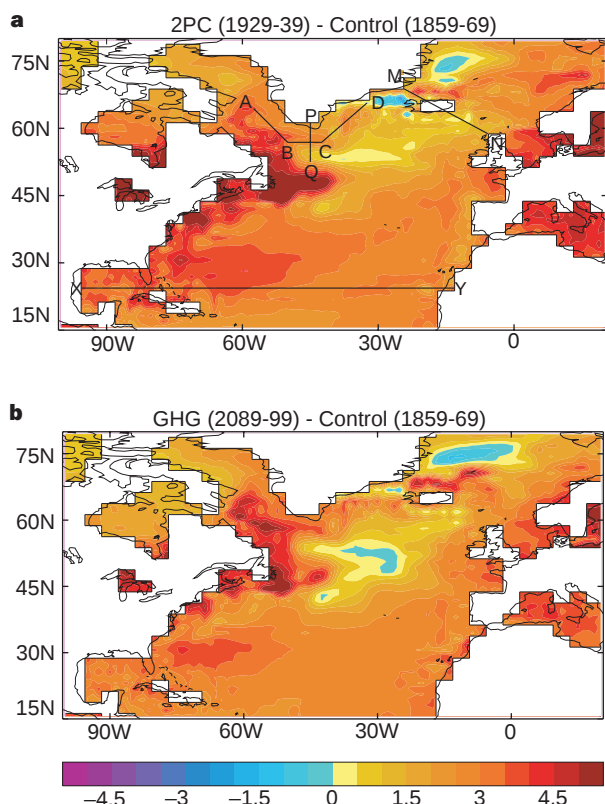


Figure 4 Changes in decadal mean sea surface temperature (°C). **a**, Difference between 2PC run 1929-39 and control 1859-69; **b**, difference between GHG run 2089-99 and control 1859-69. In **a** are also shown the approximate paths of the sections used in Fig. 1 (the GIS ridge (MN), Cape Farewell (PQ), 24° N (XY)) and Fig. 3 (ABCD).

for the response of the ocean thermohaline circulation to CO₂ forcing. If this pattern is robust, a monitoring system based on repeated hydrographic sections in the Labrador Sea and at 24° N, and current-meter measurements of the GIS overflows and the Cape Farewell boundary current, could provide a means of detection of changes in thermohaline circulation resulting from increased greenhouse-gas forcing. Much of this would build on the existing historical database^{11,21,22,29,30}. The extent to which such a signal could be detected at present depends on the natural variability in these elements of the circulation, which has not yet been fully quantified from observations. □

Received 6 November 1998; accepted 26 March 1999.

- Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1**, 841-866 (1988).
- Hall, M. M. & Bryden, H. L. Direct estimates of ocean heat transport. *Deep Sea Res.* **29**, 339-359 (1982).
- Roemmich, D. & Wunsch, C. Two transatlantic sections: meridional circulation and heat flux in the subtropical North Atlantic Ocean. *Deep Sea Res.* **32**, 619-664 (1985).
- Manabe, S. & Stouffer, R. J. Century-scale effects of increased atmospheric CO₂ on the ocean-atmosphere system. *Nature* **364**, 215-218 (1993).
- Murphy, J. M. & Mitchell, J. F. B. Transient response of the Hadley Centre coupled model to increasing carbon dioxide. Part II. Temporal and spatial evolution of patterns. *J. Clim.* **8**, 57-80 (1995).
- Cubasch, U. *et al.* Time-dependent greenhouse warming computations with a coupled ocean-atmosphere model. *Clim. Dyn.* **8**, 55-69 (1993).
- Rahmstorf, S. Risk of sea-change in the Atlantic. *Nature* **388**, 825-826 (1997).
- Houghton, J. T. *et al.* (eds) *Climate Change 1995: the Science of Climate Change* (Cambridge Univ. Press, 1996).
- Marotzke, J. & Stone, P. H. Atmospheric transports, the thermohaline circulation, and flux adjustments in a simple coupled model. *J. Phys. Oceanogr.* **25**, 1350-1364 (1995).
- Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145-149 (1995).
- Dickson, R. R. & Brown, J. The production of North Atlantic Deep Water: sources, rates and pathways. *J. Geophys. Res.* **99**, 12319-12341 (1994).
- Johns, T. C. *et al.* The second Hadley Centre coupled ocean-atmosphere GCM: model description, spinup and validation. *Clim. Dyn.* **13**, 103-134 (1997).
- Roether, W., Roussinov, V. M. & Well, R. in *Ocean Processes in Climate Dynamics: Global and Mediterranean Examples* (eds Malanotte-Rizzoli, P. & Robinson, A. R.) 371-394 (Kluwer, Dordrecht, 1994).

- Gent, P. & McWilliams, J. C. Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.* **20**, 150-155 (1990).
- Roberts, M. J. & Wood, R. A. Topographic sensitivity studies with a Bryan-Cox type ocean model. *J. Phys. Oceanogr.* **27**, 823-836 (1997).
- Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* (in the press).
- Levitus, S. & Boyer, T. P. *World Ocean Atlas 1994* (NOAA/NESDIS E/OC21, US Dept of Commerce, Washington DC, 1994).
- Houghton, J. T. *et al.* (eds) *Climate Change 1992: the Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. & Maskell, K.) (Cambridge Univ. Press, 1996).
- Weaver, A. J. & Hughes, T. M. C. On the incompatibility of ocean and atmosphere models and the need for flux adjustments. *Clim. Dyn.* **12**, 141-170 (1996).
- Bryan, F. O. Climate drift in a multi-century integration of the NCAR Climate System Model. *J. Clim.* **11**, 1455-1471 (1998).
- Bacon, S. Decadal variability in the outflow from the Nordic seas to the deep Atlantic Ocean. *Nature* **394**, 871-871 (1998).
- Clarke, R. A. Transport through the Cape Farewell-Flemish Cap section. *Rapp. P.-v. Réun. Cons. Int. Explor. Mer.* **185**, 120-130 (1984).
- Rahmstorf, S. & Ganopolski, A. Long-term global warming scenarios computed with an efficient coupled climate model. *Clim. Change* (in the press).
- IPCC Workshop Report on Rapid Non-linear Climate Change (Intergovernmental Panel on Climate Change, Bracknell, 1998).
- Dickson, R., Lazier, J., Meincke, J., Rhines, P. & Swift, J. Long-term coordinated changes in the convective activity of the North Atlantic. *Prog. Oceanogr.* **38**, 241-295 (1996).
- Stocker, T. F. & Schmittner, A. Influence of CO₂ emission rates on the stability of the thermohaline circulation. *Nature* **388**, 862-865 (1997).
- Wadley, M. R. & Bigg, G. R. Abyssal channel flow in ocean general circulation models with application to the Vema Channel. *J. Phys. Oceanogr.* **26**, 38-48 (1996).
- Doescher, R. & Redler, R. The relative importance of northern overflow and subpolar deep convection for the North Atlantic thermohaline circulation. *J. Phys. Oceanogr.* **27**, 1894-1902 (1997).
- Parilla, G., Lavin, A., Bryden, H. L., Garcia, M. & Millard, R. Rising temperatures in the subtropical North Atlantic Ocean over the past 35 years. *Nature* **369**, 48-51 (1994).
- Lazier, J. R. N. in *Natural Climate Variability on Decade-to-century Timescales* (eds Martinson, D. G. *et al.*) 295-302 (Nat. Academy Press, Washington DC, 1995).

Acknowledgements. We thank H. Cattle, G. Jenkins, J. Murphy, S. Rahmstorf, R. Stouffer, R. Thorpe and A. Weaver for comments. This work was supported by the UK Department of the Environment, Transport and the Regions.

Correspondence and requests for materials should be addressed to R.A.W. (e-mail: rwood@meto.gov.uk)

Feature-based attention influences motion processing gain in macaque visual cortex

Stefan Treue & Julio C. Martínez Trujillo

Cognitive Neuroscience Laboratory, Department of Neurology, University of Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany

Changes in neural responses based on spatial attention have been demonstrated in many areas of visual cortex¹⁻⁴, indicating that the neural correlate of attention is an enhanced response to stimuli at an attended location and reduced responses to stimuli elsewhere. Here we demonstrate non-spatial, feature-based attentional modulation of visual motion processing, and show that attention increases the gain of direction-selective neurons in visual cortical area MT without narrowing the direction-tuning curves. These findings place important constraints on the neural mechanisms of attention and we propose to unify the effects of spatial location, direction of motion and other features of the attended stimuli in a 'feature similarity gain model' of attention.

We studied the influence of attention on the sensory selectivity of neurons in visual cortex, namely direction-selective neurons in the middle temporal visual area (MT), which is important in the perception of visual motion and for motor planning^{5,6}. MT neurons have been linked directly to psychophysical performance in motion tasks⁷ and they characteristically show direction tuning curves (bell-shaped response profiles as a function of stimulus direction; Fig. 1b), which account well for psychophysical thresholds of motion perception⁸.

We recorded from neurons in area MT of two macaque monkeys while using displays of coherently moving random dot patterns (RDP) to determine what effect attention might have on these direction tuning curves. Attention might enhance the sensory gain of the neuron, that is, increase the response to all attended stimuli by the same proportion ('multiplicative modulation'), leaving the

width of the tuning curve unchanged⁹. Alternatively, attention might increase the response of a neuron only for stimuli moving in the preferred direction, thus increasing the sharpness of the neuron's tuning curve ('sharpening modulation')¹⁰.

Experiment 1 was designed to isolate the influence of spatial attention on tuning curves. One RDP was placed inside the receptive field of the neuron being recorded and the other one, moving in the same direction, was placed in the opposite visual hemifield (Fig. 1a). On a given trial, using a spatial cue, the animal's attention was directed to either one or the other stimulus, the 'target'. In both the 'attend-in' and the 'attend-out' conditions, we derived the neuron's tuning curve by randomly interleaving trials with one of 12 possible directions of movement (Fig. 1b).

Figure 1c shows a histogram of the changes in the height and width of the tuning curve between these two attentional conditions across all the cells we studied. On average, the height of the tuning curves was about 10% larger when the target was the stimulus inside the receptive field, but the tuning curves were not sharpened; instead, there was a slight, non-significant widening. The increase in the height of the tuning curve in the absence of narrowing indicates that attention has the same effect on all stimuli, that is it increases the responses by multiplicative modulation. This modulation reflects a purely spatial attentional mechanism, because the pairs of conditions compared in Fig. 1 differed only in the attended location, with the attended direction remaining the same.

Psychophysical studies suggest that attention can also be selectively allocated to stimuli that match a particular feature, without shifts in the attended location (see for example refs 11–13). To test for such effects of non-spatial, feature-based attention, we intro-

duced a variation into Experiment 1 (Fig. 2a). While the stimulus inside the receptive field now always moved in a given neuron's preferred direction, the other stimulus moved in either the same (as in the previous experiment, Fig. 2a, arrow B) or the opposite direction (Fig. 2a, arrow A). This allowed the attended direction to be switched without changing the attended location and without changing the stimulus inside the receptive field. We compared the responses when attention was directed to the stimulus outside the receptive field, moving either in the preferred or anti-preferred direction. Changing the stimulus direction outside the receptive field had no effect on the responses when that stimulus was behaviourally irrelevant, that is when the animal was attending inside the receptive field or simply fixating.

Figure 2b shows a histogram of the resulting attentional modulation across all neurons studied. Attending to the preferred motion outside the receptive field increased the response by, on average, about 13% above the response evoked when attending a null-direction stimulus outside the receptive field. This is not an effect of spatial attention, as the location of attention was unchanged between the two conditions. Rather, it represents a neural correlate of attention to stimulus feature. Comparing the responses against those evoked in trials in which none of the moving stimuli was behaviourally significant shows that this non-spatial attentional modulation is a combination of enhancement (preferred direction target, mean enhancement of ~5%) and suppression (anti-preferred direction target, mean suppression of ~6%). Thus, attending to a given direction enhances the responses of neurons whose preferred direction aligns with the attended direction and reduces the responses of those neurons preferring the opposite direction.

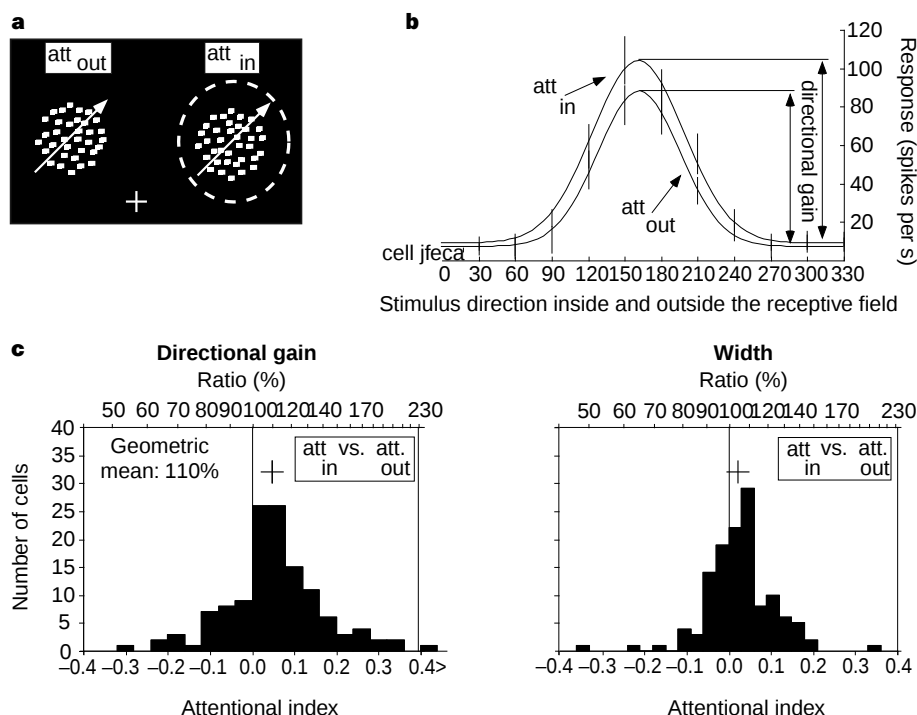


Figure 1 Experiment 1: Effect of directing attention inside versus outside the receptive field on the directional tuning curve. **a**, Sketch of the stimulus layout on the screen. One random dot pattern (RDP) was presented inside the classical receptive field (dashed circle) while the other was presented about the same distance from the fixation point in the opposite hemifield. In a given trial, both RDPs moved in the same of 12 possible directions. **b**, Examples of tuning curves. The upper curve shows the response when the monkey was attending to the stimulus inside the receptive field (marked att_{in} in **a**), and the lower curve plots the responses when the monkey was attending to the stimulus outside the receptive field (marked att_{out}). These tuning curves show an increase in directional gain and width when attention is switched from outside to inside the receptive field.

c, Histograms showing the influence of attention on the directional gain and width of the tuning curves for 131 cells. Binning is according to the attentional index $AI = (X_{in} - X_{out}) / (X_{in} + X_{out})$, where X is the gain or width in the corresponding attentional condition. The top scale shows the corresponding ratios. The left histogram shows a shift to the right, with an average AI of 0.05 (marked by the cross, where the horizontal arms span the 95% confidence interval of the mean), indicating that attention increases the height of the tuning curves on average (geometric mean) by about 10%. The right histogram shows no shift to the left, demonstrating that attention does not sharpen the tuning curves. Rather we find a small, non-significant increase in the width of the tuning curves (average increase: 4%, $P > 0.05$ in paired t -test).

This influence is far reaching; our stimuli were as much as 20° apart and in opposite visual hemifields.

Having demonstrated attentional modulation of about equal size with shifts in the spatial location of attention and with feature-based effects in the absence of a shift of the attended location we plot the combined effect of the two modulations. Figure 2c compares responses when the animal was attending the anti-preferred stimulus outside the receptive field with those trials when attention was directed into the receptive field to the stimulus moving in the preferred direction (Fig. 2a, arrows A and C). The attentional modulation (25% on average) is the sum of the shifts shown in Figs 1c and 2b, emphasizing that feature-based attentional effects can be additively combined with modulations based on the spatial location of attention. Comparing the two attentional conditions against responses when neither of the stimuli was behaviourally significant shows that the attentional modulation is a combination of the suppressive effect of switching attention to the null direction outside the receptive field (~6% suppression) and the enhancing effect of directing attention into the receptive field onto the preferred direction (~15% enhancement).

Previous studies demonstrated a strong response modulation when attention was switched between stimuli that were both inside the receptive field^{11–14}. In our third experiment we tested whether the absence of attentional sharpening of the tuning curves persists under these circumstances by placing two stimuli side-by-side inside the receptive field. Pattern A always moved in the anti-preferred direction of the cell. To generate a tuning curve, pattern B moved in one of twelve directions of motion. Again, in a given trial, either one of the patterns was designated as the target. By plotting the response of the neuron as a function of the direction of motion of pattern B, a tuning curve could be determined for each of the two attentional conditions. Figure 3 shows an example of these tuning curves for one cell together with the 'sensory' tuning curve, recorded when neither of the two patterns inside the receptive field was behaviourally relevant.

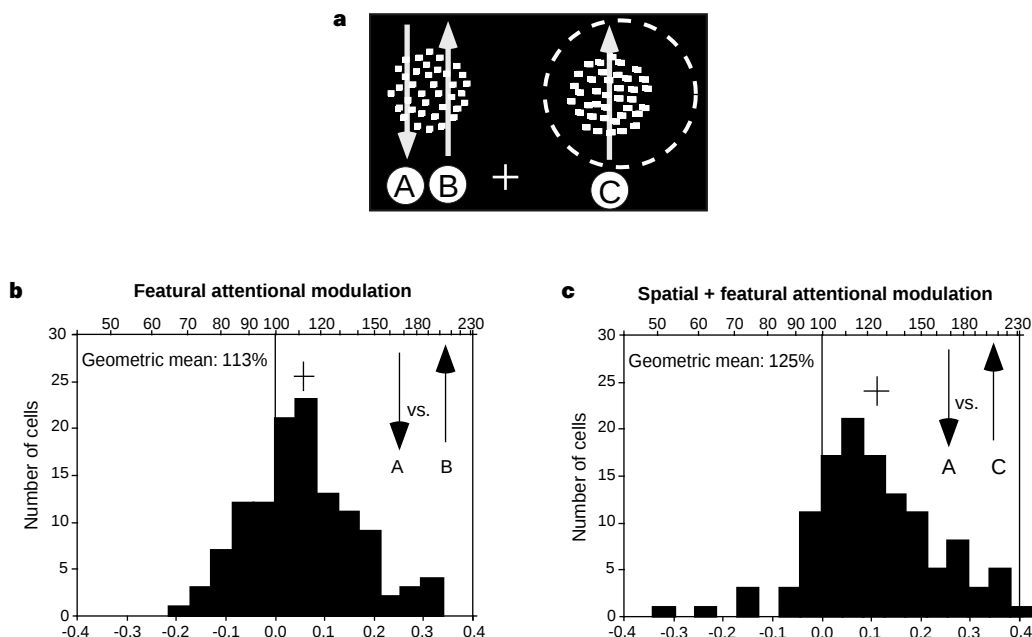


Figure 2 Experiment 2: Non-spatial effects of attention and the summing of spatial and featural attentional modulation. **a**, Stimulus conditions used. The stimulus inside the receptive field always moved in the cell's preferred direction (upward pointing arrow); the stimulus outside moved in either the same (B) or the opposite direction (A). Trials in which the animal was instructed to attend to A, B and C were presented in an interleaved fashion. **b**, Histogram comparing responses of 131 cells when attention was on the preferred (B) or anti-preferred (A) direction

The lower panels of Fig. 3 show histograms of the attentional modulation of the directional gain and tuning width. As in Experiment 1, attention increases the directional gain of the neuron, although now with a mean effect of about 60%. Even with these very strong response modulations, no sharpening of the tuning was observed.

Our results demonstrate a physiological correlate of non-spatial, feature-based attention by showing response modulations in the absence of spatial shifts of attention. We further show that spatial and feature-based attention represent summable processes that have a multiplicative effect on the responses of neurons. Such attentional modulations resemble changes to a neuron's sensory gain and thus can be mimicked by sensory effects, such as reducing the luminance contrast of a stimulus, which similarly does not change the tuning width of direction-selective neurons^{14,15}, suggesting that response modulation based on attentional and sensory aspects employ common mechanisms.

Non-spatial, feature-based modulation of sensory responses has been observed in imaging studies^{16,17} and using psychophysical paradigms^{12–14}. However, previous studies did not show an unambiguous single-cell correlate of this effect, because they investigated attentional selection based on stimulus features, leaving open the possibility that the modulation itself is based on stimulus location^{18,19}, or confounded a change in the attended feature with a simultaneous change in attended location⁹.

Although the absence of a sharpening of the tuning curves is in contrast to one report from areas in the ventral visual pathway¹⁰, it closely matches another⁹, indicating that attention may work in similar ways in the dorsal and ventral visual pathways. A recent study attempting to model psychophysical orientation discrimination performance in dual-task attentional paradigms has indicated that the observed performance can only be accounted for by models that implement sharpening of tuning curves with attention²⁰. As we have found no indication for such sharpening, further studies will be necessary to understand the reasons for this discrepancy. The

outside the receptive field. The histogram is shifted to the right (mean shift 13%) indicating an increased response when the stimulus moved in the cell's preferred direction. **c**, Histogram comparing responses when attention was on anti-preferred motion outside (A) or the stimulus inside the receptive field (C). The histogram is shifted to the right, indicating an increased response when the target was inside the receptive field.

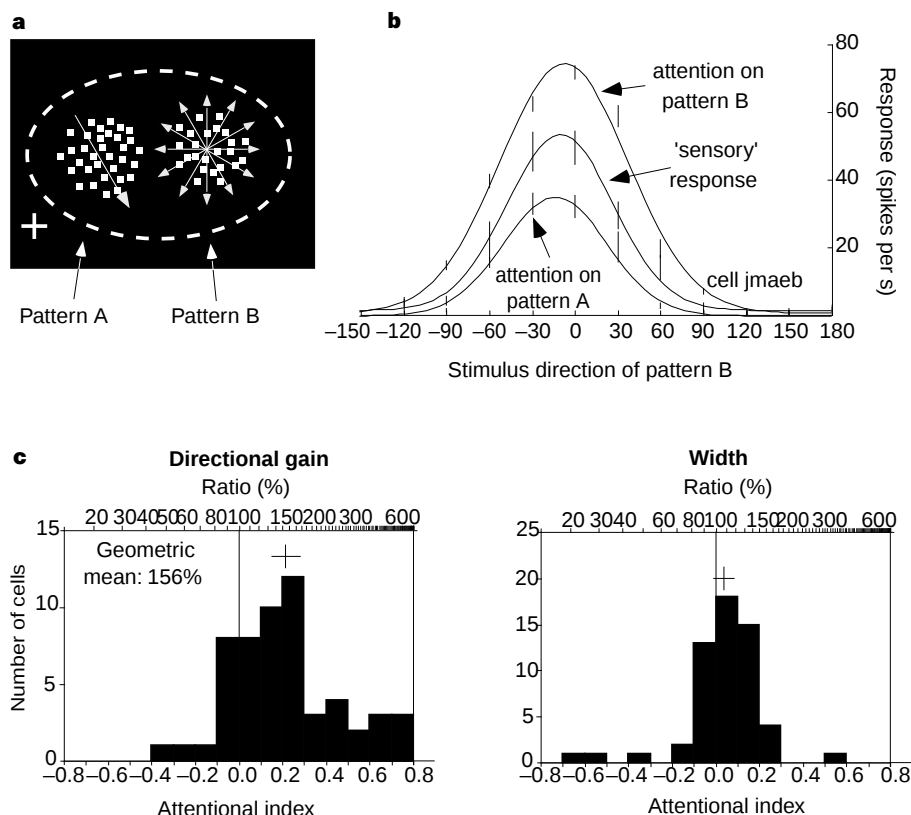


Figure 3 Experiment 3: Effect of directing attention to one of two stimuli inside the receptive field. **a**, Stimulus configurations. Both patterns were presented inside the receptive field. Pattern A always moved in the cell's anti-preferred direction, pattern B in one of 12 possible directions. **b**, Tuning curves when pattern B was the target (upper curve), when pattern A was the target (lower curve) and when neither pattern was behaviourally relevant (central curve) because the animal was instructed to respond to a luminance change at the fixation point. **c**,

Histograms of the attentional modulation of the tuning curve across 56 cells. The mean increase in directional gain is about 60% (which is a combination of response enhancement when switching attention from the 'sensory' condition to pattern B and of suppression when switching to pattern A). Again there is no narrowing of the tuning width. On average, width is increased by 8% (non-significant, $P > 0.1$).

attentional enhancement we observe does support better stimulus discriminability even without tuning sharpening, by increasing the slope of the tuning curve^{21,22}.

A 'biased competition model' has been proposed²³, which supposes that attention influences the competition between two stimuli for access to a given cell in favour of the attended stimulus. This is achieved by increasing the strength of the signal coming from the population of input cells activated by the attended stimulus²⁴. The attentional modulation we observed when attention was switched between two stimuli inside the receptive field (Experiment 3) conforms to the predictions of this model. Given the model's emphasis on competition within the receptive field, our findings in Experiment 2 (attentional effects outside the receptive field and the additivity of spatial and non-spatial effects when attention is switched into the receptive field) do not seem to be predicted by the model. We propose that spatial and non-spatial attentional effects can be unified in a 'feature similarity gain model', in which the up- or downregulation of the gain of a sensory neuron reflects the similarity of the features of the currently behaviourally relevant target and the sensory selectivity of the neuron along all target dimensions. Thus this up- or downregulation will also affect neurons whose receptive fields do not include the attended stimulus location. The relevant target features include the spatial location, direction of motion and presumably others. Correspondingly, the sensory selectivity of the neuron includes the location of its receptive field (or of the smaller receptive fields of its input neurons), its preferred direction of motion and presumably other preferred features. This model not only provides a good account of

other physiological studies of attention that included conditions without competing stimuli inside the receptive field^{2,9}, but also incorporates the idea, from psychophysical and imaging studies as well as other modelling attempts, that non-spatial stimulus features can be the basis of attentional effects.

Methods

Cells and recording. Our recording methods have been described elsewhere^{3,25}. All animal procedures complied with the NIH Guide for Care and Use of Laboratory Animals and were approved by the local animal care committee. Cells were determined to be from MT by their physiological characteristics (directionality and receptive field position and size) as well as by the position of the electrode in the cortex. We also recorded from MST cells, finding similar but larger attentional modulations to those reported here.

Stimuli. Random dot patterns were created by plotting bright dots at a density of 5 dots per deg square within a circular stationary virtual aperture on a dark computer monitor. Dots moved coherently at the preferred speed of the neurons and were re-plotted to the opposite side when they crossed the border of the aperture. The size of the aperture was chosen so that the stimulus did not exceed the boundaries of the classical receptive field.

Trials. Every trial started with the appearance of the fixation cross. After it was foveated, a stationary RDP (the cue) signalled the location of the 'target'. The monkey then depressed a lever. 200–300 ms later the cue disappeared (Experiments 1 and 2) or moved for 400 ms (Experiment 3). After a blank interval of 65 ms (Experiments 1 and 2) or 270 ms (Experiment 3) two moving RDPs appeared, one (the 'target') at the location of the cue and one (the 'distractor') at another location. The monkey's task was to release the lever when the target changed speed or direction (which occurred at a random point

between 270 and 4,000 ms after target onset) and to ignore changes in the distractor. Failure to respond within a reaction-time window, responding to a change in the distractor or deviating the gaze (monitored with a scleral search coil) by more than 1° from the fixation point caused the trial to be aborted without reward. The change in the target and distractors was selected so as to be challenging for the animal. In experiments 1 and 2 the animal correctly completed, on average, 79% of the trials, broke fixation in 11%, might have responded to the distractor stimulus in 6% and responded too early or not at all in 5% of the trials. In Experiment 3 the corresponding values are 78, 13%, 8% and 2%. In none of the three experiments was there a difference between the performances for the two possible targets. Differences between average eye positions during trials where one or the other stimulus was the target were very small, with only an average shift of 0.02° in the direction of the shift of position between the stimuli. Only correctly completed trials were considered. Firing rates were determined by computing the average neuronal response across trials for 1,000 ms starting 200 ms after the beginning of the target stimulus movement.

Tuning curves. Tuning curves were derived by fitting the responses to the 12 directions presented with gaussian functions: $r_{\text{null}} + \text{dirGain} \times \exp(-0.5(\text{dir} - \text{prefdir})^2 / \text{width}^2)$. The four parameters of a gaussian curve capture the four features of a direction-selective cell: preferred direction (*prefdir*), response to the anti-preferred direction (*r_{null}*), the directional gain (*dirGain*; the maximal response modulation) and the selectivity or tuning width (*width*; the range of directions the neuron responds to).

Received 14 December 1998; accepted 14 April 1999.

1. Moran, J. & Desimone, R. Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784 (1985).
2. Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70**, 909–919 (1993).
3. Treue, S. & Maunsell, J. H. R. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382**, 539–541 (1996).
4. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
5. Newsome, W. T., Wurtz, R. H., Dürsteler, M. R. & Mikami, A. Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *J. Neurosci.* **5**, 825–840 (1985).
6. Newsome, W. T. & Paré, E. B. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–2211 (1988).
7. Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebriani, S. & Movshon, J. A. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* **13**, 87–100 (1996).
8. Snowden, R. J., Treue, S. & Andersen, R. A. The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Exp. Brain Res.* **88**, 389–400 (1992).
9. McAdams, C. J. & Maunsell, J. H. R. Effects of attention on orientation tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
10. Spitzer, H., Desimone, R. & Moran, J. Increased attention enhances both behavioral and neuronal performance. *Science* **240**, 338–340 (1988).
11. Lankheet, M. J. M. & Verstraten, F. A. J. Attentional modulation of adaptation to two-component transparent motion. *Vision Res.* **35**, 1401–1412 (1995).
12. Duncan, J. & Nimmo-Smith, I. Objects and attributes in divided attention: Surface and boundary systems. *Percept. Psychophys.* **58**, 1076–1084 (1996).
13. Valdes-Sosa, M., Bobes, M. A., Rodriguez, V. & Pinilla, T. Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *J. Cogn. Neurosci.* **10**, 137–151 (1998).
14. Reynolds, J. H. & Desimone, R. Attention and contrast have similar effects on competitive interactions in macaque area V4. *Soc. Neurosci. Abstr.* **23**, 302 (1997).
15. Treue, S. & Martinez, J. C. Attentional modulation of direction-selective responses in MT/MST resembles the effect of reducing contrast of unattended stimuli. *Soc. Neurosci. Abstr.* **24**, 1249 (1998).
16. O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A. & Savoy, R. L. Voluntary attention modulates fMRI activity in human MT-MST. *Neuron* **18**, 591–598 (1997).
17. Beauchamp, M. S., Cox, R. W. & DeYoe, E. A. Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J. Neurophysiol.* **78**, 516–520 (1997).
18. Motter, B. C. Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J. Neurosci.* **14**, 2178–2189 (1994).
19. Chelazzi, L., Miller, E. K., Duncan, J. & Desimone, R. A neural basis for visual search in inferior temporal cortex. *Nature* **363**, 345–347 (1993).
20. Itti, L., Braun, J., Lee, D. K. & Koch, C. Attentional modulation of human pattern discrimination psychophysics reproduced by a quantitative model. *Neural Information Processing Systems* (in the press).
21. McAdams, C. J. & Maunsell, J. H. R. Attention enhances neuronal responses without altering orientation selectivity in macaque area V4. *Neurosci. Abstr.* **22**, 1197 (1996).
22. Martinez, J. & Treue, S. Attention does not sharpen direction-tuning curves in macaque monkey MT/MST neurons. *Soc. Neurosci. Abstr.* **24**, 649 (1998).
23. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
24. Reynolds, J. H., Chelazzi, L. & Desimone, R. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.* **19**, 1730–1753 (1999).
25. Treue, S. & Maunsell, J. H. R. Effects of attention on the processing of motion in macaque visual cortical areas MT and MST. *J. Neurosci.* submitted.

Acknowledgements. This work was supported by the MWF-Württemberg. J.C.M. is a fellow of the Graduiertenkolleg Neurobiologie, Tübingen.

Poleward shifts in geographical ranges of butterfly species associated with regional warming

Camille Parmesan^{*,†}, Nils Ryrholm[‡], Constantí Stefanescu[§], Jane K. Hill^{||}, Chris D. Thomas[¶], Henri Descimon[#], Brian Huntley^{||}, Lauri Kaila[☆], Jaakko Kullberg[☆], Toomas Tammaru^{**}, W. John Tennent^{††}, Jeremy A. Thomas^{‡‡} & Martin Warren^{§§}

^{*} National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101, USA

[‡] Evolutionary Biology Centre, Section of Zoological Ecology, Uppsala University, Norbyvägen 18 D, S-752 36 Uppsala, Sweden

[§] Butterfly Monitoring Scheme, Can Liro, 08458 Sant Pere de Vilamajor, Barcelona, Spain

^{||} Environmental Research Centre, Department of Biological Sciences, University of Durham, Durham DH1 3LE, UK

[¶] Centre for Biodiversity and Conservation, School of Biology, University of Leeds, Leeds LS2 9JT, UK

[#] Laboratoire de Systématique Évolutive, Université de Provence, 3 place Victor Hugo, 13331 Marseille, Cedex 3, France

[☆] Finnish Museum of Natural History, Division of Entomology, University of Helsinki, P.O. Box 17, Helsinki FIN-00014, Finland

^{**} Institute of Zoology and Botany, Estonian Agricultural University, Riia 181, EE-51014 Tartu, Estonia

^{††} Biogeography and Conservation Laboratory, The Natural History Museum (BMNH), London SW7 5BD, UK

^{‡‡} Furzebrook Research Station, Institute of Terrestrial Ecology, Wareham, Dorset BH20 5AS, UK

^{§§} Butterfly Conservation, P.O. Box 444, Wareham, Dorset BH20 5YA, UK

Mean global temperatures have risen this century, and further warming is predicted to continue for the next 50–100 years^{1–3}. Some migratory species can respond rapidly to yearly climate variation by altering the timing or destination of migration⁴, but most wildlife is sedentary and so is incapable of such a rapid response. For these species, responses to the warming trend should be slower, reflected in poleward shifts of the range. Such changes in distribution would occur at the level of the population, stemming not from changes in the pattern of individuals' movements, but from changes in the ratios of extinctions to colonizations at the northern and southern boundaries of the range. A northward range shift therefore occurs when there is net extinction at the southern boundary or net colonization at the northern boundary. However, previous evidence has been limited to a single species⁵ or to only a portion of the species' range^{6,7}. Here we provide the first large-scale evidence of poleward shifts in entire species' ranges. In a sample of 35 non-migratory European butterflies, 63% have ranges that have shifted to the north by 35–240 km during this century, and only 3% have shifted to the south.

We analysed distributional changes broadly spread over the past century for non-migratory species of butterfly whose northern boundaries were in northern Europe and whose southern boundaries were in southern Europe or northern Africa. We excluded some data where circumstances suggested that range boundaries were controlled or altered by non-climatic factors. This yielded a subset of sufficient quality for us to detect distributional changes predicted by models of global warming, yet is unbiased with respect to such changes. However, because data for some species were excluded at either their northern or southern boundaries, we

[†] Present address: Integrative Biology, Patterson Laboratories, University of Texas, Austin, Texas 78712, USA.

present three analyses: (1) northern boundary analysis, including all species with suitable data for the northern boundary; (2) southern boundary analysis, including all species with suitable data for the southern boundary; (3) whole-range analysis, encompassing species with suitable data for both boundaries, covering the intersection of datasets 1 and 2.

The northern boundary analysis used data for 52 species, of which the northern boundaries have extended northwards in the past 30–100 years for 65%, been stable for 34%, and retracted southwards for 2% (Table 1; binomial test, 34 extended north, 1 retracted south, $P \ll 0.001$). There is a high level of agreement among countries. Of the 34 species extending northwards, 24 have data from more than one country; for 71% of these, data from two or more countries indicate northward extension. Disagreements occur where one country classifies a species as stable while another country classifies it as extending its range. In no case did a species show conflicting trends, being classified as extending in one country and retracting in another. A detailed history of range extension in Great Britain since 1915 is shown for *Pararge aegeria* in Fig. 1.

The southern boundary analysis used data for 40 species, of which the southern boundaries have retracted northwards in the past 30–100 years for 22%, remained stable for 72%, and extended southwards for 5% (Table 2, binomial test, 9 retracting north, 2 extending south, $P < 0.04$). The proportion of stable boundaries is significantly higher along southern than along northern boundaries (Fisher's exact test, 72% versus 34%, $P < 0.001$). This could be because southern boundaries may be determined by factors other than temperature, such as moisture, or by non-climatic factors, such as competition⁸. Alternatively, the difference may be because southern study regions are more mountainous, and populations may have shifted elevation or slope aspect, rather than latitude. Finally, southern study areas may have experienced less climatic change than northern areas. Further data are required to test the relative merits of these hypotheses.

The whole-range analysis encompassed 35 species with data at both northern and southern boundaries, of which 63% shifted northwards, 29% were stable at both boundaries, 6% shifted southwards, and 3% extended at both boundaries (Table 3, Figs 2–4; binomial test, 22 moving north, 2 moving south, $P \ll 0.001$). These patterns do not exhibit any obvious phylogenetic component. All four families or subfamilies that are represented in the analysis by at least three species include both shifting and non-shifting species.

We conducted analyses only on boundary changes, not on distributional changes in non-boundary regions. For example, although *Carterocephalus palaemon* has become extinct in the southern parts of both Great Britain and Finland (possibly as a result of habitat alterations), its northern boundary in each of those countries is stable, as is its southern boundary in the Pyrenees (Catalonia), so it is categorized as non-shifting (Fig. 1).

Nearly all northward shifts involved extensions at the northern boundary with the southern boundary remaining stable (about two-thirds of 22 species; see Fig. 3), or retracting northwards (about one-third of 22 species). Thus, most species effectively expanded the size of their range when shifting northwards. The remaining species shifted their entire distribution northwards, more or less without altering the range size. For example, *Heodes tityrus* has completely disappeared from the Montseny region of Catalonia, where it was easily seen early this century. At the northern limit of the same species, there were only three isolated records in Estonia before 1998, when several breeding populations were found (Fig. 4).

The magnitudes of these range shifts (35–240 km along a single boundary) are on the order of 5–50 times the colonization distances achieved by comparable butterflies in single colonization events^{9,10}. Northward extensions therefore seem to have been a consequence of sequential establishment of new populations, each giving rise to further colonizations. In *P. aegeria*, at least one (and usually several) new population has been established in each of the 10-km grid squares to the north of its historical distribution (Fig. 1).

Table 1 Northern boundary assessment

Family/subfamily*	Extended northwards	Stable	Retracted southwards
Papilionidae	<i>Parnassius apollo</i> <i>Parnassius mnemosyne</i>		
Pieridae	<i>Anthracaris cardamines</i> <i>Aporia crataegi</i>	<i>Gonepteryx rhamni</i>	
Lycaenidae	<i>Agrodiaetus amanda</i> <i>Everes argiades</i> † <i>Glaucopsyche alexis</i> <i>Heodes tityrus</i> <i>Quercusia quercus</i> <i>Strymonidia pruni</i> <i>Hamearis lucina</i> (Riodininae)	<i>Celastrina argiolus</i> <i>Cyaniris semiargus</i> <i>Heodes alciphron</i> <i>Heodes virgaureae</i> <i>Plebejus argus</i> <i>Strymonidia w-album</i> <i>Thecla betulae</i>	
Nymphalidae	<i>Apatura iris</i> <i>Araschnia levana</i> <i>Argynnis paphia</i> <i>Argynnis niobe</i> <i>Clossiana dia</i> <i>Brenthis daphne</i> <i>Inachis io</i> † <i>Limenitis camilla</i> <i>Limenitis populi</i> <i>Polygonum c-album</i>	<i>Argynnis adippe</i> <i>Brenthis ino</i> <i>Clossiana selene</i> <i>Melitaea cinxia</i>	<i>Apatura ilia</i>
Satyrinae	<i>Aphantopus hyperantus</i> <i>Coenonympha glycerion</i> <i>Hipparchia semele</i> <i>Lasiommata megera</i> <i>Lopinga achine</i> <i>Maniola jurtina</i> <i>Melanargia galathea</i> <i>Pararge aegeria</i> <i>Pyronia tithonus</i>	<i>Coenonympha arcania</i> <i>Erebia aethiops</i> <i>Lasiommata maera</i>	
Hesperiidae	<i>Erynnis tages</i> <i>Ochlodes venatus</i> <i>Thymelicus lineola</i> <i>Thymelicus sylvestris</i>	<i>Carterocephalus palaemon</i> <i>Thymelicus acteon</i>	

Data are from one or more countries: Great Britain, Sweden, Finland and Estonia.

* Taxonomy for all species are from ref. 19.

† Highly mobile species included because the distance of boundary change is much greater than individual dispersal distances.

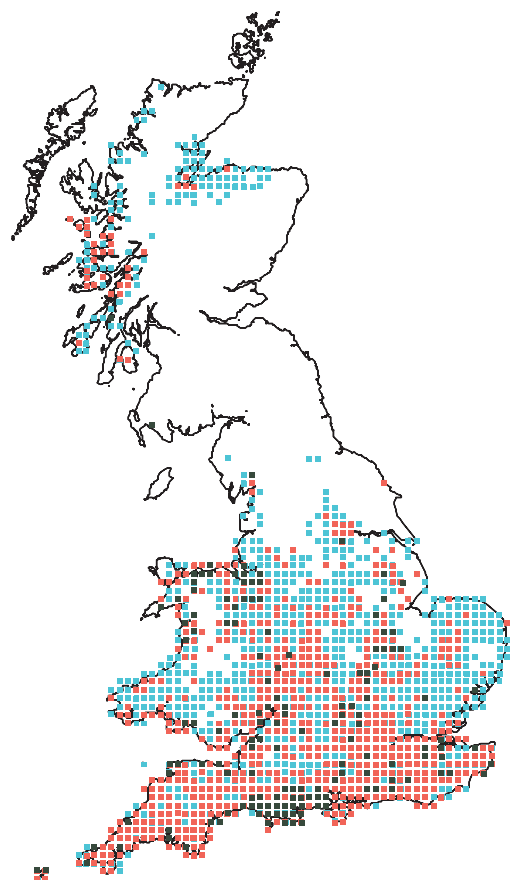


Figure 1 Twentieth-century changes in the range of *Pararge aegeria* in Great Britain, plotted by presence in Ordnance Survey 10 × 10 km grid squares. A coloured grid cell indicates more than one population in 1915–1939 (black), 1940–1969 (red) or 1970–1997 (blue).

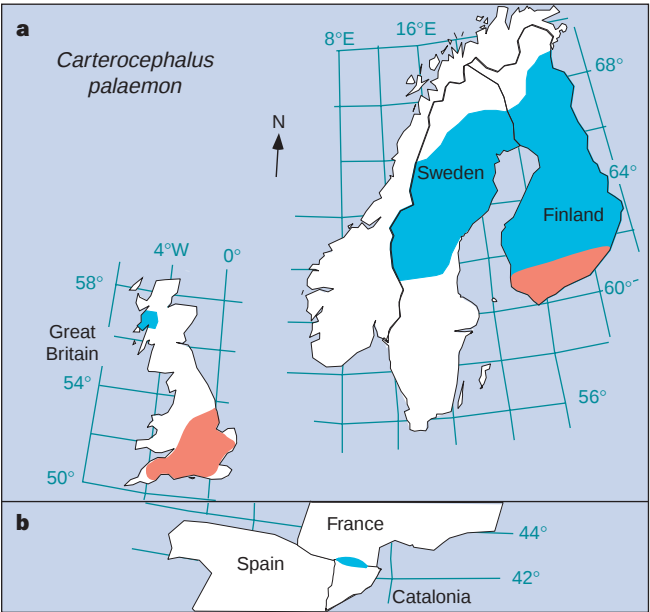


Figure 2 Distribution of a non-shifting species, *Carterocephalus palaemon*. Orange indicates areas of retraction, where all historical populations are extinct. Blue indicates stable areas, with no change. **a**, Northern boundary; **b**, southern boundary.

Changing land use is unlikely to account for these boundary changes. We selected species least likely to have been affected by modern habitat alterations. For Great Britain, we tested whether this selection might have biased our conclusion. We reanalysed the data to incorporate all 38 non-migratory species that have northern boundaries within Great Britain, including those known to be severely habitat-restricted or to have suffered severe habitat loss. We still found 47% with northward extension and only 8% with

Table 2 Southern boundary assessment

Family/subfamily	Retracted northwards	Stable	Extended southwards
Papilionidae	<i>Parnassius apollo</i> <i>Parnassius mnemosyne</i>		
Pieridae		<i>Aporia crataegi</i>	
Lycaenidae	<i>Everes argiades</i> <i>Glaucopsyche alexis</i> <i>Heodes tityrus</i> <i>Hamearis lucina</i> (Riodininae)	<i>Agrodiaetus amanda</i> <i>Celastrina argiolus</i> <i>Cupido minimus</i> <i>Heodes alciphron</i> <i>Heodes virgaureae</i> <i>Quercusia quercus</i> <i>Strymonidia w-album</i> <i>Thecla betulae</i>	
Nymphalinae	<i>Clossiana dia</i> <i>Clossiana selene</i>	<i>Apatura ilia</i> <i>Apatura iris</i> <i>Argynnis adippe</i> <i>Argynnis paphia</i> <i>Brenthis daphne</i> <i>Clossiana euphrosyne</i> <i>Limenitis camilla</i> <i>Limenitis populi</i> <i>Limenitis reducta</i> <i>Polygonum c-album</i> *	<i>Brenthis ino</i> <i>Araschnia levana</i>
Satyrinae	<i>Aphantopus hyperantus</i> <i>Minois dryas</i>	<i>Coenonympha arcania</i> <i>Lasiommata maera</i> <i>Maniola jurtina</i> <i>Melanargia galathea</i> <i>Nechipparchia statilinus</i> <i>Pyronia tithonus</i>	
Hesperiidae		<i>Carterocephalus palaemon</i> <i>Thymelicus acteon</i> † <i>Thymelicus sylvestris</i>	

Data are from one or more countries: France, Catalonia (Spain), Morocco, Algeria and Tunisia.
* Disappeared from 6 out of 11 historically recorded sites in Africa, but the southernmost site is still present; the conservative assessment is to call this 'stable'.
† Disappeared from 3 out of 6 historically recorded sites in Africa, but the southernmost site is still present; the conservative assessment is to call this 'stable'.

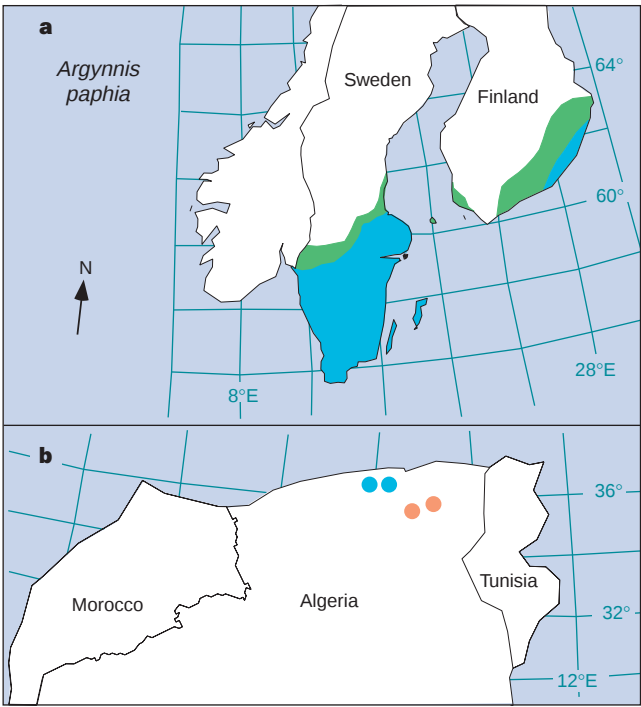


Figure 3 A northward-shifting species, *Argynnis paphia*, stable at its southern boundary and extending at its northern edge. **a**, Northern boundary; for Fennoscandia, blue represents the distribution in 1970, green in 1997. *A. paphia* is present in Great Britain but this is excluded from the analysis because of habitat loss with recent woodland management changes. **b**, Southern boundary; for northern Africa, each historical site (recorded from 1906 to 1912) is shown as a separate circle. Blue indicates that the species is present in the current census; orange indicates absence in the current census.

southward retraction in the past 30 years (binomial test, 18 extended north, 3 retracted south, $P < 0.001$). Several of these species have extended northwards across heavily cultivated landscapes that are clearly less suitable for those species than they were a hundred years ago. In our study area as a whole, habitat loss has been greater in the northern countries than in the southern ones¹¹. Although habitat loss has been severe in northern Africa, we controlled for this effect by only including sites where the habitat was still suitable for butterflies, regardless of whether the species of

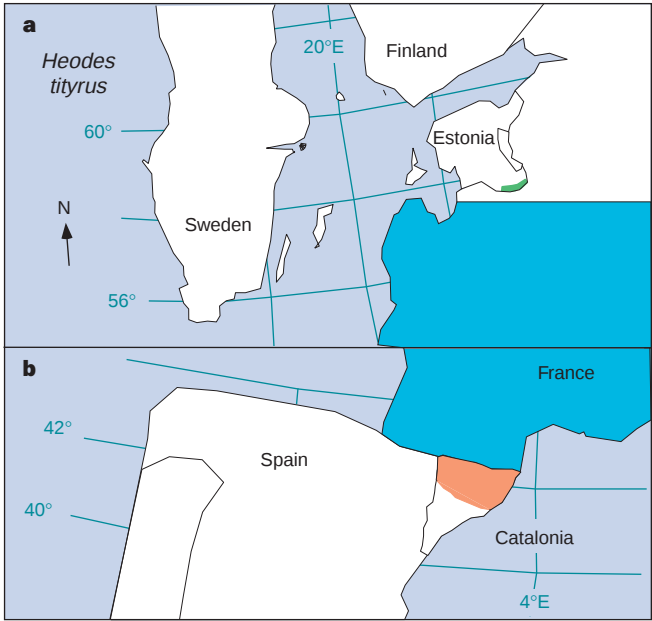


Figure 4 A northward-shifting species, *Heodes tityrus*, retracting at its southern boundary and extending at its northern boundary. **a** Northern boundary. Blue indicates the distribution for most of this century, taken from published maps²⁰, green indicates the distribution in 1998. There are also three other records in Estonia, once each in 1942, 1994 and 1996. **b**, Southern boundary. Orange indicates multiple populations recorded in the first half of this century, all now extinct.

interest was present in the recent census. Thus, the influence of human land-use changes on our data set should tend to generate southward range shifts. Our conclusions are therefore conservative with respect to these effects.

Europe has warmed by about 0.8 °C this century, shifting the climatic isotherms northwards by an average of 120 km (ref. 3), on the same order of magnitude as the detected range shifts. In western North America, isotherms have shifted north by 105 km (ref. 12), while the mean location of populations of the butterfly *Euphydryas*

Table 3 Whole-range assessment

	No shift			Southward shift			Northward shift		
Northern limit:	Stable	Retraction	Extension	Stable	Retraction	Retraction	Extension	Stable	Extension
Southern limit:	Stable	Retraction	Extension	Stable	Stable	Extension	Stable	Retraction	Retraction
Papilionidae									<i>P. apollo</i> <i>P. mnemosyne</i>
Pieridae							<i>A. crataegi</i>		
Lycaenidae	<i>C. argiolus</i> <i>H. alciphron</i> <i>H. virgaureae</i> <i>S. w-album</i> <i>T. betulae</i>						<i>A. amanda</i> <i>Q. quercus</i> <i>H. lucina</i>		<i>E. argiades</i> <i>G. alexis</i> <i>H. tityrus</i>
Nymphalinae	<i>A. adippe</i>		<i>A. levana</i>	<i>B. ino</i>	<i>A. ilia</i>		<i>A. iris</i> <i>A. paphia</i> <i>B. daphne</i> <i>L. camilla</i> <i>L. populi</i> <i>P. c-album*</i>	<i>C. selene</i>	<i>C. dia</i>
Satyrinae	<i>C. arcania</i> <i>L. maera</i>						<i>M. galathea</i> <i>M. jurtina</i> <i>P. tithonus</i>		<i>A. hyperantus</i>
Hesperiidae	<i>C. palaemon</i> <i>T. acteon*</i>						<i>T. sylvestris</i>		

* Large density decreases in south: *T. acteon* extinct at 3 of 6 sites; *P. c-album* extinct at 6 of 11 sites.

editha shifted north by 92 km (ref. 5). Consistency across taxa and continents indicates that butterfly species in the northern hemisphere are shifting generally northwards in response to a common environmental change.

Nearly all population-dynamic studies have concluded that butterflies, and insects in general, are sensitive to temperature^{13–16}. Although the correlational nature of our study limits our ability to determine causal factors, the summed knowledge of butterfly biology, including numerous experimental studies, implies that the northward shifts represent responses to increased temperatures. If this is correct, it has implications for climate-sensitive organisms with similar population structures characterized by discrete populations and limited dispersal, such as many beetles, grasshoppers, rodents and frogs.

Given the relatively slight warming in this century compared with predicted increases of 2.1 to 4.6 °C for the next century², our data indicate that future climate warming could become a major force in shifting species' distributions. But it remains to be seen how many species will be able to extend their northern range margins substantially across the highly fragmented landscapes of northern Europe. This could prove difficult for all but the most efficient colonizers. □

Methods

Assembling data sets. Data were assembled and analysed for non-migratory butterflies with northern boundaries in Great Britain, Sweden, Finland or Estonia, and southern boundaries in southeastern France, Catalonia (in Spain), Algeria, Tunisia or Morocco. A species was excluded if: (1) it is extremely habitat-restricted, requiring for example such a narrow combination of microclimate, plant phenology and other characteristics that it is highly localized, even within habitats containing its host; or (2) it cannot tolerate even a modest level of human-mediated habitat modification. A particular country's data for a species were excluded if, within that country: (1) the species' range is limited by host-plant distribution; (2) the range boundary lies in an area with so little potential habitat that it could have changed because of habitat alteration; or (3) the species suffered severe habitat loss. In some northern regions, specialists on dry meadows, calcareous grasslands and marshes were excluded owing to severe declines in these habitats in recent decades. In northern Africa, data from sites that no longer provided good butterfly habitat during the period of late censuses were excluded; if good habitat was still available, irrespective of the presence or absence of the target species, the data were included. (Species lists for each country are available on request from C.P.)

Data quality. The data are not of consistent resolution or temporal span within or among countries. Data for northern Africa are from historically famous collection sites (mainly from 1901 to 1932) that were re-censused by W.J.T. in the 1990s. Catalonia had a period of extensive collecting from 1900 to 1930, with moderate records since 1950. Distributions for the past 20 years are well documented by C.S. and a network of modern recorders; high-quality data collection began five years ago through a systematic, government-funded monitoring scheme. For France and Estonia, H.D. and T.T. had carried out field research and worked with naturalist societies. Final assessments incorporated literature records and combined data from amateur recorders. For most species, the changes analysed represent the past 20 years for Estonia and 40 years for southeastern France; a few species could be tracked from the 1920s and 1930s. In the northern third of Finland, Sweden and Great Britain (north Scotland and Lapland), collecting and recording activity was relatively poor for the first half of the century, so, except for the most conspicuous species, we focused on changes since the 1950s. However, in the southern two-thirds of those countries, observations have been detailed throughout the century, yielding data on a decadal scale. For detailed descriptions of data sets for each country, see Supplementary Information.

Assessment within a country. A species' boundary was defined as stable where census data show that populations recorded from earlier in this century are still present; it was defined as extended where recent censuses found new populations in areas outside the known historical distribution that had been visited by earlier collectors or recorders; and it was defined as retracted where historically recorded populations that defined a boundary are absent in recent

censuses whose sampling effort was sufficient for absence of a record to indicate true absence of the species. Where population changes at a range boundary were small, we conservatively assessed that species as stable at that boundary. Thus, extended or retracted distributions involved boundary changes of at least several individual dispersal distances and could not be the result of a few ephemeral populations or be an artefact of stochastic sampling.

Assessment among countries, along a boundary. A single status for each boundary (northern or southern) was determined by the following: a boundary is deemed stable if it shows no changes in any country along the same boundary; a boundary is retracting if it has moved towards the centre of the distribution within one or more countries and is stable in other countries along the same boundary; a boundary is extending if it has moved away from the centre of the distribution within one or more countries and is stable in other countries along the same boundary.

Assessment of range status. Where the status of both northern and southern boundaries has been determined for a species, a northward shift is defined as either both boundaries moving north, or one moving north and the other remaining stable; a southern shift is similarly defined.

Analysis of fluctuations. This methodology leaves unclear how to classify species whose boundaries have fluctuated over time, with periods of both extension and retraction. For the southern boundaries, such information is not available; the same is true for most of the northern boundaries. However, for some species in northern regions, we have sufficient data to show northward extensions in the 1930s and 1940s and southward retractions in the 1950s and 1960s^{16–18}. These fluctuations coincide with decadal temperature fluctuations in northern Europe³ that are overlaid on the background temperature rise over the century. Species showing these fluctuations have all re-extended their northern boundaries concurrent with the recent warming trend (starting in 1970)¹⁸, and all have current distributions that are further north than they were at the beginning of the century. Therefore, we have recorded them as 'extended'.

Received 23 November 1998; accepted 12 April 1999.

1. Easterling, D. R. *et al.* Maximum and minimum temperature trends for the globe. *Science* **277**, 364–367 (1997).
2. Intergovernmental Panel on Climate Change. *Climate Change 1995. Report of Working Group I* (eds Houghton, J. T. *et al.*) (Cambridge Univ. Press, 1996).
3. Beniston, M. *et al.* in *The Regional Impacts of Climate Change. Intergovernmental Panel on Climate Change Working Group II* (eds Watson, R. T., Zinyowera, M. C. & Moss, R. H.) 149–185 (Cambridge Univ. Press, 1998).
4. Root, T. L. in *Elements of Change 1996* (eds Hassol, S. J. & Katzenberger, J.) 203–206 (Aspen Global Change Institute, Aspen, Colorado, 1997).
5. Parmesan, C. Climate and species' range. *Nature* **382**, 765–766 (1996).
6. Barry, J. P., Baxter, C. H. K., Sagarin, R. D. & Gilman, S. E. Climate-related long-term faunal changes in a California rocky intertidal community. *Science* **267**, 672–675 (1995).
7. Grabherr, G., Gottfried, M. & Pauli, H. Climate effects on mountain plants. *Nature* **369**, 448 (1994).
8. Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B. & Wood, S. Making mistakes when predicting shifts in species range in response to global warming. *Nature* **391**, 783–786 (1998).
9. Thomas, C. D. & Hanski, I. in *Metapopulation Biology: Ecology, Genetics and Evolution* (eds Hanski, I. & Gilpin, M. E.) 359–386 (Academic, London, 1997).
10. Thomas, C. D., Thomas, J. A. & Warren, M. S. Distributions of occupied and vacant butterfly habitats in fragmented landscapes. *Oecologia* **92**, 563–567 (1992).
11. Baldock, D. Agriculture and habitat loss in Europe (CAP Discussion paper No 3, Gland, Switzerland, 1990).
12. Karl, T. R., Knight, R. W., Easterling, D. R. & Quayle, R. G. Indices of climate change for the United States. *Bull. Am. Meteorol. Soc.* **77**, 279–292 (1996).
13. Dennis, R. L. H. *Butterflies and Climate Change* (Manchester Univ. Press, 1993).
14. Coope, G. R. in *Extinction Rates* (eds Lawton, J. & May, R.) 55–74 (Oxford Univ. Press, 1995).
15. Uvarov, B. P. Insects and climate. *Trans. Entomol. Soc. Lond.* **79**, 174–186 (1931).
16. Ford, E. B. *Butterflies* (Collins, London, 1945).
17. Kaisila, J. *Immigration und Expansion der Lepidopteren in Finnland in den Jahren 1869–1960* (ACTA Entomologica Fennica, Helsinki, 1962).
18. Pollard, E. & Eversham, B. C. in *Ecology and Conservation of Butterflies* (ed. Pullin, A. S.) 23–36 (Chapman & Hall, London, 1995).
19. Ackery, P. R. Systematic and faunistic studies on butterflies. In *Biology of Butterflies* (eds Vane-Wright, R. I. & Ackery, P. R.) 9–21 (Academic, London, 1984).
20. Tolman, T. *Butterflies of Britain and Europe* (HarperCollins, London, 1997).

Supplementary information is available on Nature's World-Wide Web site (www.nature.com) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank the huge number of amateur lepidopterists throughout Europe who have collected most of the data. Data sets are from private collectors' records, regional lists and publications, the Swedish Museum of Natural History, the Lepidopterological Society of Sweden, the Natural History Museum, London (BMNH), the Museum of Zoology of Barcelona, Societat Catalana de Lepidopterologia and Butterfly Monitoring Scheme (Departament de Medi Ambient, Generalitat de Catalunya), Biological Records Centre (ITE, Monks Wood, UK), Butterfly Conservation (UK), Estonian Naturalists' Society, Lepidopterological Society of Finland, and Finnish Museum of Natural History. This project was facilitated by the National Center for Ecological Analysis and Synthesis, P. R. Ackery, D. Blakeley and M. C. Singer. We thank A. N. Cohen, T. Lewinsohn, F. Micheli, W. Porter, J. Roughgarden, M. C. Singer, F. Wagner, R. I. Vane-Wright and M. Willig for comments on the manuscript.

Correspondence and requests for materials should be addressed to C.P. (e-mail: parmesan@nceas.ucsb.edu).

Assignment of circadian function for the *Neurospora* clock gene *frequency*

Martha Merrow^{*}, Michael Brunner[†] & Till Roenneberg^{*}

Institutes for ^{*} Medical Psychology and [†] Physiological Chemistry, Ludwig Maximilians University, Goethestrasse 31-33, 80336 München, Germany

Circadian clocks consist of three elements: entrainment pathways (inputs), the mechanism generating the rhythmicity (oscillator), and the output pathways that control the circadian rhythms. It is difficult to assign molecular clock components to any one of these elements. Experiments show that inputs can be circadianly regulated¹⁻³ and outputs can feed back on the oscillator^{4,5}. Mathematical simulations indicate that under- or overexpression of a gene product can result in arrhythmicity, whether the protein is part of the oscillator or substantially part of a rhythmically expressed input pathway⁶. To distinguish between these two possibilities, we used traditional circadian entrainment protocols^{7,8} on a genetic model system, *Neurospora crassa*.

Rather than being directly driven by environmental changes, the *zeitgeber*, circadian rhythms are controlled by endogenous oscillators, and synchrony with *zeitgeber* is achieved by entrainment of the oscillator. Unlike simple reactions to external signals, the complex mechanisms of entrainment reflect the robust momentum of the running clock as well as a time-of-day-specific responsiveness to *zeitgeber* signals. As a result, circadian clocks adopt specific stable phase relationships to *zeitgeber* cycles. In short cycles, entrained rhythms lag relative to the *zeitgeber*; in long cycles, they lead. When the cycle length is approximately half that of the endogenous free-running period (FRP, measured in constant conditions) then circadian rhythms (unlike driven rhythms) often 'frequency demultiply'⁹ (skip a cycle). The *Neurospora* clock is monitored by observing rhythmic conidiation (asexual spore formation) while it is growing in glass tubes. Null mutants of the *Neurospora* clock gene *frequency* (*frq*) are arrhythmic in constant conditions, induction of the protein FRQ resets the circadian phase, and FRQ is self-regulated by negative feedback. These findings are theoretically

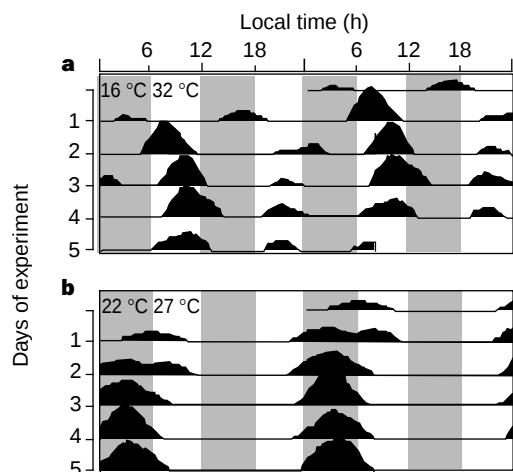


Figure 1 Circadian entrainment of *frq*⁺ by 12 h temperature cycles (**a**, 16–32°C; **b**, 22–27°C; 6 h low (grey) and 6 h high (white) temperature in constant darkness). Conidiation (black areas) is double plotted: the data from 2 days are graphed in a continuum, starting at the top left of the graph with day 1, and moving down by one day on each line of the vertical axis. Double plotting the data in this way allows trends to be visualized across midnight. Note that conidiation occurs in warmth in high-amplitude cycles and in cold in low-amplitude cycles.

consistent with FRQ being either a rhythmically expressed input component⁶ or an oscillator component¹⁰⁻¹². Entraining conditions could distinguish between these possibilities. If FRQ is not essential for the oscillator, then both FRQ-sufficient (*frq*⁺, wild-type, FRP = 22 h; *frq*¹, short-period mutant, FRP = 16 h; *frq*⁷, long-period mutant, FRP = 29 h) and FRQ-deficient strains (*frq*⁹, frame shift by base-pair deletion resulting in premature stop codon; the phenotype is indistinguishable from the null mutant, *frq*¹⁰)¹³⁻¹⁶ should show typical circadian entrainment.

To test whether temperature *zeitgeber* cycles entrain or merely drive the conidiation rhythm, short cycles (6:6) were imposed on *frq*⁺. In high-amplitude cycles (16 to 32°C), a driven component is present once each cycle (every 12 h, Fig. 1a). Circadian regulation is, however, apparent from the asymmetric conidiation in alternate cycles. In low-amplitude cycles (22 to 27°C, Fig. 1b), the driven component disappears, and conidiation occurs every other cycle (once every 24 h; a frequency demultiplication). Phase changes due to different *zeitgeber* amplitudes (compare Fig. 1a and b) are typical of entrainment of circadian clocks⁷. Thus, temperature cycles result in true circadian entrainment of *Neurospora* and do not merely drive conidiation.

We compared the different *frq* mutants in temperature cycles of different lengths (see typical examples in Fig. 2a and b). All strains tested, including the FRQ-deficient mutant *frq*⁹, establish a stable phase relationship to the *zeitgeber* cycle, although *frq*¹ and *frq*⁹ are almost 180° out of phase. The two rhythmic strains, *frq*¹ and *frq*⁷ (FRPs of 16 and 29 h), also entrain in anti-phase (Fig. 2c). The different phase position of each strain over a wide range of *zeitgeber*

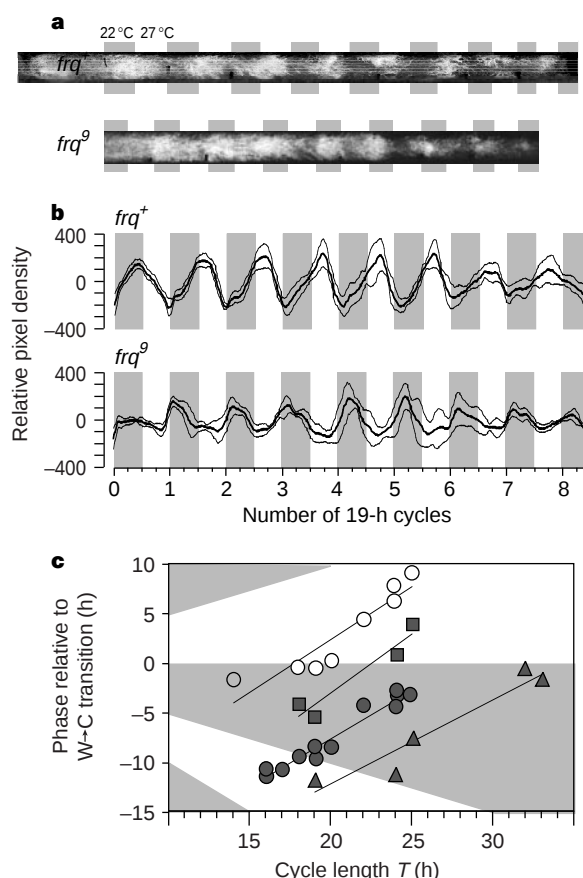


Figure 2 Temperature cycles and circadian entrainment of *frq* mutants. **a**, *frq*⁺ and *frq*⁹ in 19 h (9.5:9.5) temperature cycles (22–27°C, shading as in Fig. 1). **b**, Average pixel densities of 6 tubes (thick line) for the experiment shown in **a** (±s.d. indicated by thin lines). **c**, Phase plots: Onsets of conidiation for *frq*⁹ (open circles), *frq*¹ (squares), *frq*⁺ (filled circles) and *frq*⁷ (triangles) are expressed in real hours ('phase') relative to the warm to cold (W → C) transition. Negative values indicate that onsets lag the temperature transition. Lines represent linear fits.

periods (T) depends on the strain's FRP (judged by those which are rhythmic in constant conditions). When T equals the FRP, conidiation starts 5–6 h (by extrapolation) after the warm-to-cold transition. As expected, conidiation leads in longer and lags in shorter cycles. This typical circadian entrainment^{7,17} produces the series of parallel lines. The phase plot of the arrhythmic strain (frq^9 , open circles) shows the same slope as the rhythmic strains (filled symbols), indicating that it is controlled by the same temperature-entrainable oscillator. Its relative position in Fig. 2c indicates a virtual FRP of about 12 h (by extrapolation, frq^9 conidiates 5–6 h into the cold when $T = 12$). Characteristic circadian entrainment is a property of all tested strains and, therefore, does not require functional FRQ protein. The light-blind *white collar* mutants^{18,19}, which are also arrhythmic in constant conditions, have a mutation in a gene that is required for frq transcription. Their responses to temperature cycles are superimposable with those of frq^9 (data not shown), probably because of the lack of FRQ. FRQ appears to determine FRP and hence also the phase position in temperature cycles.

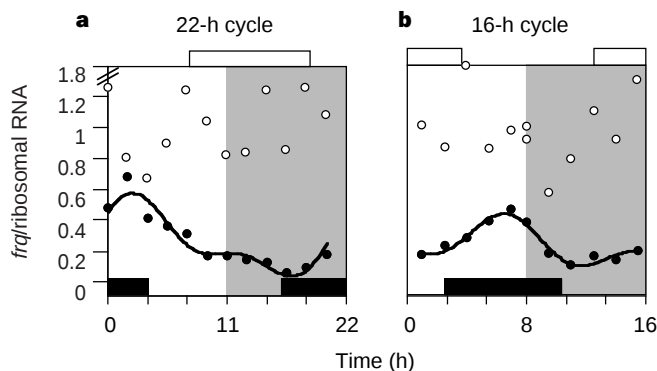


Figure 3 frq mRNA levels in temperature cycles (22–27 °C, shading as in Fig. 1) for frq^+ (filled circles) and frq^9 (open circles). **a**, 22-h cycle; **b**, 16-h cycle. Rectangles at the top and bottom of the graphs indicate the timing of the entrained conidial bands in glass tubes (filled, frq^+ ; open, frq^9). Curves represent significant 2-harmonic-cosine fits ($r = 0.95$ for the 22-h cycle and 0.97 for the 16-h cycle). Curve fits for frq^9 were not significant ($r = 0.62$ and 0.60 , respectively).

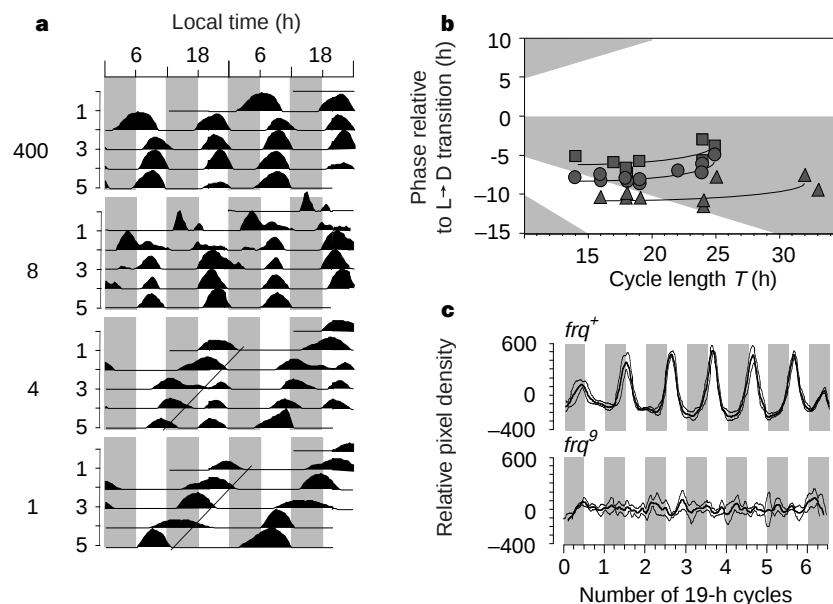


Figure 4 Light:dark cycles do not produce circadian entrainment. **a**, 12-h light:dark cycles (dark phase in grey; fluences in $\text{nE m}^{-2} \text{s}^{-1}$ at left; constant temperature 22 °C). **b**, Phase plots for light cycles (400 $\text{nE m}^{-2} \text{s}^{-1}$, symbols as in Fig. 2c). Onsets of conidiation are related to the light-dark (L $\rightarrow$ D) transition, expressed in real

In free-running cultures of FRQ-sufficient strains, the overall abundance of frq messenger RNA oscillates with a periodicity that matches conidiation^{10,11}. mRNA accumulation in different *zeitgeber* periods (compare phase angles in Fig. 3a and b) also mirrors conidiation and shows that frq is controlled by the oscillator and not merely by temperature²⁰. In contrast, RNA levels in frq^9 are high and non-systematically variable throughout the temperature cycle, as they are in free-running cultures¹⁰. Thus, the temperature-entrainable oscillator requires FRQ protein to control frq accumulation (by negative feedback^{10,21}), but not to control rhythmic conidiation.

To compare light and temperature as *zeitgeber* signals for *Neurospora*, we used the demultiplication protocol described above. Light:dark cycles (6:6, 400 $\text{nE m}^{-2} \text{s}^{-1}$) drive conidiation in every cycle (Fig. 4a, top panel). The fluence rate was, therefore, titrated down to find conditions where light:dark cycles induce circadian entrainment as shown by frequency demultiplication and systematic phase-angle changes due to different *zeitgeber* amplitudes. At as low as 8 $\text{nE m}^{-2} \text{s}^{-1}$, conidiation is still driven every 12 h (Fig. 4a). At 4 $\text{nE m}^{-2} \text{s}^{-1}$ (equivalent to the light of a night with a full moon), a free-running rhythm appears in addition to the light-driven component. At 1 $\text{nE m}^{-2} \text{s}^{-1}$, conidiation runs free as in constant darkness. Irrespective of *zeitgeber* amplitude, neither frequency demultiplication nor phase angle changes occur in full photoperiod cycles, as they do in temperature cycles (Fig. 1). A third characteristic of circadian entrainment, phase angle changes due to different *zeitgeber* periods, is also absent in light:dark cycles (Fig. 4b). Onset of conidiation is independent of T , occurring a set amount of time after the light:dark transition (equal to 1/3 of the strain's FRP; compare with Fig. 2c). The lack of frequency demultiplication and the fixed phase angles show that full photoperiods drive, rather than entrain, conidiation of the FRQ-sufficient strains (frq^+ , frq^1 , frq^7). The threshold level above which light drives conidiation is the same as the level that renders *Neurospora*²² and *Drosophila pseudoobscura*²³ arrhythmic in constant light.

Molecular and physiological light responsiveness *per se* is not impaired in frq -null strains^{12,13,24}, but the temperature-entrainable oscillator is unresponsive to light:dark cycles in frq^9 (Fig. 4c; see also ref. 25) and frq^{10} (data not shown). We predict, therefore, that FRQ is part of a clock-regulated light input pathway to the temperature-

hours. **c**, frq^+ and frq^9 in a 19-h light:dark cycle (constant temperature 26 °C, 400 $\text{nE m}^{-2} \text{s}^{-1}$ white light). Statistical analysis revealed no rhythmicity in frq^9 . Experiments at 22 °C gave the same result.

entrainable oscillator. Note that the different rhythmic *frq* mutants show strain-specific responses to short light pulses^{25,26}.

In laboratory experiments, the effect of light on the *Neurospora* clock, even at moonlight levels, is so strong that it apparently stops the oscillator²². In nature, such a clock would be no more than an hourglass timer, functioning only during the darkest nights. The light-cycle experiments shown here were done in constant temperature, but when light and temperature are given in anti-phase in a 24 h cycle (cold with light and warm with dark)²⁰, temperature dominates as the *zeitgeber*. When *frq*⁺ is exposed to temperature cycles (22–27 °C) in constant light (which results in arrhythmicity in constant temperature), conidiation is robustly rhythmic (data not shown). Thus, temperature cycles 'gate' light transduction, enabling the *Neurospora* clock to continue throughout the day. Our results show that FRQ is unnecessary for entraining an oscillator with generic circadian properties in temperature cycles. However, without FRQ, the circadian clock cannot synchronize to light:dark cycles and self-sustained rhythmicity is almost never seen, except under special conditions^{13,14,27}. As part of a circadianly regulated light input pathway^{1–3}, FRQ apparently supplies the clock with sufficient amplitude for self-sustained rhythmicity (*frq*⁰ remains arrhythmic after release from temperature entrainment) and sets the period in the circadian range. This role of FRQ is especially relevant in view of the recent reports of clock-regulated light transduction (via cryptochrome) to the clock in flies and mammals^{1,28,29}. □

Methods

Strains and media. *bd*, *bd frq*¹ and *bd frq*⁰ are standard lab strains (provided by the Dunlap Lab, Dartmouth Medical School). *bd frq*⁷ was obtained from the Fungal Genetics Stock Center. Solid media for glass tubes contained 1× Vogels salts¹⁴, 0.3% glucose, 0.5% L-arginine, 2% agar and 10 ng biotin per ml. Liquid media was the same, except for 0.5% glucose and the lack of agar.

Zeitgeber cycles. Depending on the *zeitgeber*, half of each cycle was spent in low temperature or darkness, the other half in high temperature or light (Lumilux Interna, Osram). Temperature cycles were created in custom-made incubators. For increases in temperature, 90% of the end point (22 to 27 °C) was attained in 48 min. For decreases (27 to 22 °C) 108 min was required for 90% of the change. Light cycles were administered at constant temperature (26 or 22 °C). Fluence was titrated with neutral density filters (Rosco). Glass tubes were inoculated and incubated for about 1 day in constant light and ambient lab temperature (23 ± 1 °C) before transfer to the *zeitgeber* cycle (low temperature or darkness).

mRNA analysis. Glass tubes and liquid cultures (3.7 × 10⁸ conidia per 10 ml of media, in a 50 ml Erlenmeyer flask) were simultaneously inoculated and transferred (see above) to the experimental set-up. Samples were collected over the course of the third full cycle, by which time conidiation rhythms are generally stably entrained. RNA was prepared and analysed by standard methods^{12,21}. Loading differences were normalized by relating *frq* to ribosomal RNA. For each individual data set, the average quantification in the *frq*⁰ series is 1 and all other values (*frq*⁺ and *frq*⁰) are expressed as proportions thereof.

Data analysis. Images of glass tubes were digitized with an Apple Color One scanner, stored as PICT files and analysed with CHRONO³⁰. Conidiation was quantified by the number of white pixels in each vertical line of the image and expressed as deviation around the non-rhythmic trend. 3–6 tubes were analysed for each time series and variable. Onsets of conidiation were defined as upward transition through the non-rhythmic trend (see zero lines in Figs 2b and 4c) of the averaged time series.

Received 12 January; accepted 19 April 1999.

- Emery, P., So, W. V., Kaneko, M., Hall, J. C. & Rosbash, M. CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* **95**, 669–679 (1998).
- Roenneberg, T. & Deng, T.-S. Photobiology of the *Gonyaulax* circadian system: I Different phase response curves for red and blue light. *Planta* **202**, 494–501 (1997).
- Roenneberg, T. & Foster, R. G. Twilight times—light and the circadian system. *Photochem. Photobiol.* **66**, 549–561 (1997).
- Block, G., Geusz, M., Khalsa, S., Michel, S. & Whitmore, D. (eds). *Cellular Analysis of a Molluscan Retinal Biological Clock* 51–66 (Wiley, Chichester, 1995).
- Mrosovsky, N., Salmon, P. A., Menaker, M. & Ralph, M. R. Nonphotic phase shifting in hamster clock mutants. *J. Biol. Rhythms* **7**, 41–50 (1992).

- Roenneberg, T. & Merrow, M. Molecular circadian oscillators—an alternative hypothesis. *J. Biol. Rhythms* **13**, 167–179 (1998).
- Aschoff, J. (ed.) *Biological Rhythms* (Plenum, New York, 1981).
- Aschoff, J. in *Proc. XIII Int. Conf. Soc. Chronobiol.* (Pavia 1977) (Halberg, F. et al., eds) 105–112 (Il Ponte, Milan, 1981).
- Bruce, V. Environmental entrainment of circadian rhythms. *Cold Spring Harbor Symp. Quant. Biol.* **25**, 29–48 (1960).
- Aronson, B. D., Johnson, K. A., Loros, J. J. & Dunlap, J. C. Negative feedback defining a circadian clock: autoregulation of the clock gene frequency. *Science* **263**, 1578–1584 (1994).
- Garceau, N. Y., Liu, Y., Loros, J. J. & Dunlap, J. Alternative initiation of translation and time specific phosphorylation yield multiple forms of essential clock protein FREQUENCY. *Cell* **89**, 469–476 (1997).
- Crosthwaite, S. K., Loros, J. J. & Dunlap, J. C. Light-induced resetting of a circadian clock is mediated by a rapid increase in frequency transcript. *Cell* **81**, 1003–1012 (1995).
- Loros, J. J. & Feldman, J. F. Loss of temperature compensation of circadian period length in the *frq*-9 mutant of *Neurospora crassa*. *J. Biol. Rhythms* **1**, 187–198 (1986).
- Aronson, B. D., Johnson, K. A. & Dunlap, J. C. The circadian clock locus frequency: a single ORF defines period length and temperature compensation. *Proc. Natl Acad. Sci. USA* **91**, 7683–7687 (1994).
- Feldman, J. F. & Hoyle, M. N. Isolation of circadian clock mutants of *Neurospora crassa*. *Genetics* **75**, 605–613 (1973).
- Merrow, M. W. & Dunlap, J. C. Intergeneric complementation of a circadian rhythmicity defect: phylogenetic conservation of structure and function of the clock gene frequency. *EMBO J.* **13**, 2257–2266 (1994).
- Somers, D. E., Webb, A. A., Pearson, M. & Kay, S. A. The short-period mutant, *toc1-1*, alters circadian clock regulation of multiple outputs throughout development in *Arabidopsis thaliana*. *Development* **125**, 485–494 (1998).
- Crosthwaite, S. K., Dunlap, J. C. & Loros, J. J. *Neurospora wc-1* and *wc-2*: Transcription, photo-responses, and the origin of circadian rhythmicity. *Science* **276**, 763–769 (1997).
- Russo, V. E. Blue light induces circadian rhythms in the *bd* mutant of *Neurospora*: double mutants *bd,wc-1* and *bd,wc-2* are blind. *J. Photochem. Photobiol. B* **2**, 59–65 (1988).
- Liu, Y., Merrow, M., Loros, J. J. & Dunlap, J. C. How temperature changes reset a circadian oscillator. *Science* **281**, 825–829 (1998).
- Merrow, M., Garceau, N. & Dunlap, J. Dissection of a circadian oscillation into discrete domains. *Proc. Natl Acad. Sci. USA* **94**, 3877–3882 (1997).
- Sargent, M. L., Briggs, W. R. & Woodward, D. O. Circadian nature of a rhythm expressed by an invertaseless strain of *Neurospora crassa*. *Plant Physiol.* **41**, 1343–1349 (1956).
- Winfrey, A. T. Suppressing *Drosophila* circadian rhythm with dim light. *Science* **183**, 970–972 (1974).
- Arpaia, G., Loros, J. J., Dunlap, J. C., Morelli, G. & Macino, G. The interplay of light and the circadian clock. *Plant Physiol.* **102**, 1299–1305 (1993).
- Chang, B. & Nakashima, H. Effects of light-dark cycles on the circadian conidiation rhythm in *Neurospora crassa*. *J. Plant Res.* **110**, 449–453 (1997).
- Dharmananda, S. *Studies of the Circadian Clock of Neurospora crassa: Light-Induced Phase Shifting* (University of California, Santa Cruz, 1980).
- Lakin-Thomas, P. L. Choline depletion, *frq* mutations, and temperature compensation of the circadian rhythm in *Neurospora crassa*. *J. Biol. Rhythms* **13**, 268–277 (1998).
- Thresher, R. J. et al. Role of mouse cryptochrome blue-light photoreceptor in circadian photo-responses. *Science* **282**, 1490–1494 (1998).
- Stanewsky, R. et al. The *cry*² mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell* **95**, 681–692 (1998).
- Roenneberg, T. & Taylor, W. Automated Recordings of Bioluminescence—with special reference to the analysis of circadian rhythms. *Methods Enzymol.* **305** (in the press).

Acknowledgements. This work is dedicated to the memory of J. Aschoff. We thank V. Schiewe, A. Kohnert and G. Meyer for excellent technical assistance; the members of the Woody Hastings lab for comments on the manuscript; and the DFG, the Meyer-Struckmann-Stiftung and the Münchner Medizinische Wochenschrift for support.

Correspondence and requests for materials should be addressed to T.R. (e-mail: till.roenneberg@imp.med.uni-muenchen.de) or M.M. (e-mail: martha@imp.med.uni-muenchen.de).

Control of organ shape by a secreted metalloprotease in the nematode *Caenorhabditis elegans*

Robert Blelloch* & Judith Kimble†‡§

* Program in Cell and Molecular Biology, † Howard Hughes Medical Institute, and ‡ Departments of Biochemistry and Medical Genetics and Laboratory of Molecular Biology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

The molecular controls governing organ shape are poorly understood. In the nematode *Caenorhabditis elegans*, the gonad acquires a U-shape by the directed migration of a specialized 'leader' cell, which is located at the tip of the growing gonadal 'arm'. The *gon-1* gene is essential for gonadal morphogenesis: in *gon-1* mutants, no arm elongation occurs and somatic gonadal structures are severely malformed². Here we report that *gon-1* encodes a secreted protein with a metalloprotease domain and

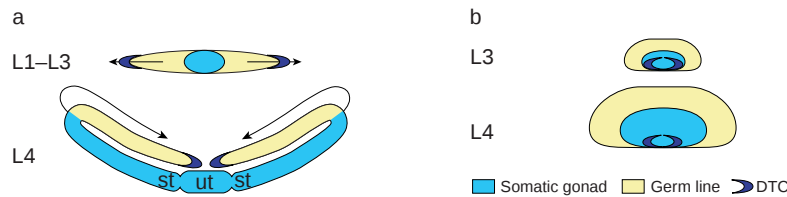


Figure 1 The *gon-1* gene is required for morphogenesis of the gonad. L1-L4, first to fourth larval stages. Arrows, direction of arm extension. **a**, Wild-type hermaphrodite. The gonad possesses two U-shaped ovotestes extending symmetrically from central gonadal structures that include uterus (ut), sper-

matheca (st) and sheath (sh). Hermaphrodite leader cells, called distal tip cells (DTCs) lead arm extension (for a review, see ref. 23). **b**, The *gon-1* gene is required for gonadal shape. In *gon-1* mutants, arm extension does not occur, and the gonad develops as a disorganized mass of somatic and germline tissues².

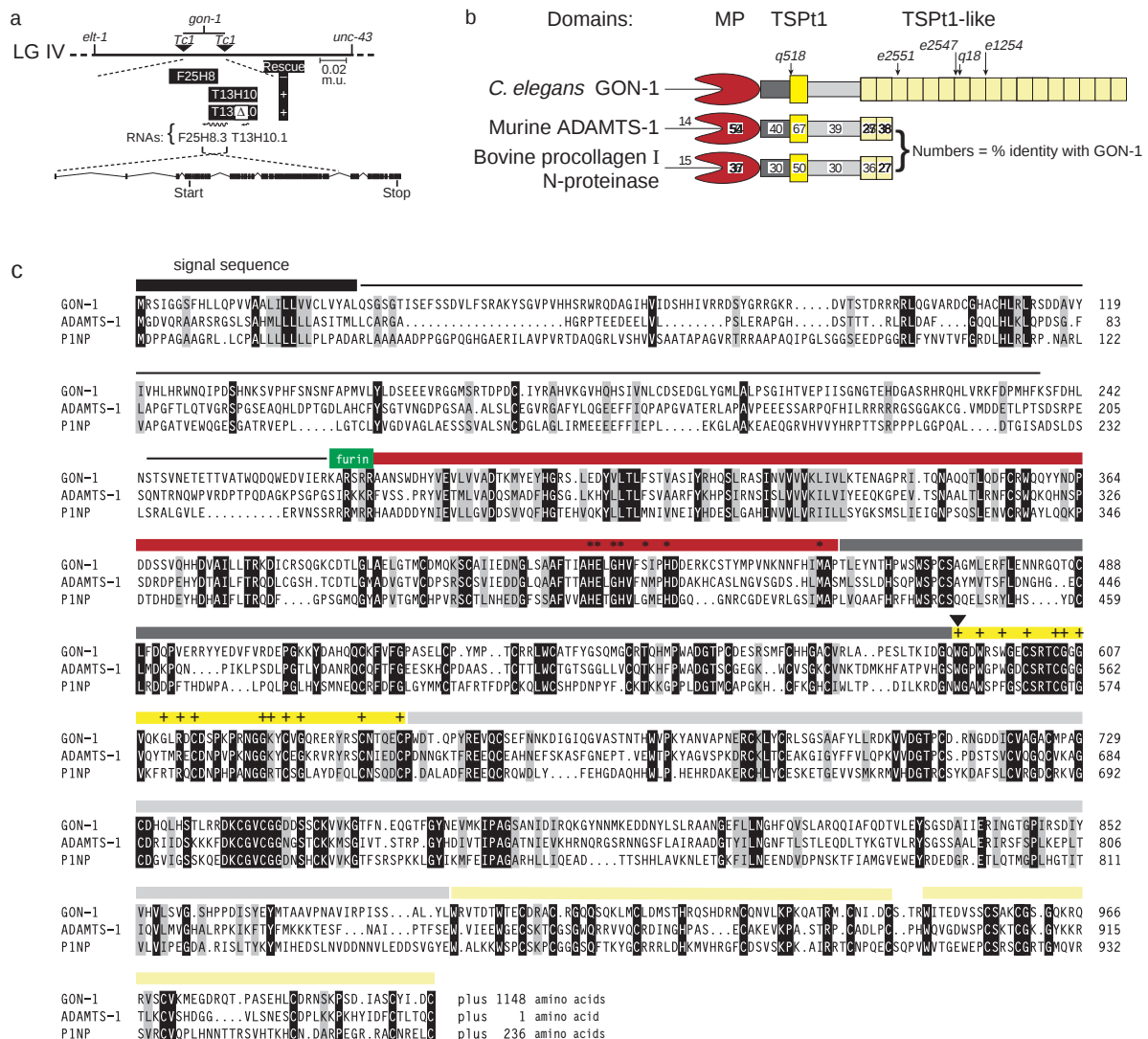


Figure 2 The *gon-1* gene and protein. **a**, Molecular cloning. Top, genetic map: *gon-1* maps to LG IV between *Tc1* insertions pKP5151 and pKP614. Middle, *gon-1* transcript F25H8.3 spans two cosimids (F25H8 and T13H10); T13H10 rescues *gon-1(q518)*, as does a deleted version that removes the smaller transcript T13H10.1. Below, the *gon-1* transcript, deduced by comparing genomic and cDNA sequences²⁴. Black boxes, exons; translation start and stop sites are indicated. **b**, Domains of GON-1 and other family members. Prodomain, thin line; MP, metalloprotease (red); conserved cysteine-rich regions (CR1, dark grey) and CR2 (light grey); TSPt1, thrombospondin type 1 motif (bright yellow); TSPt1-like motifs, amino acids 889-1923 (light yellow). Arrows, *gon-1* mutations: *q518* (TGG → TGA at amino acid 591), *e2551* (TGG → TAG at amino acid 1,069), *e2547*

(TGG → TGA at amino acid 1,229), *q18* (TGG → TAG at amino acid 1,234) W → stop; *e1254* (CGA → TGA at amino acid 1,345) R → stop. **c**, GON-1 sequence alignment with two MPT family members. Colour coding as in **b**; black bar, predicted signal sequence; green bar, predicted furin-cleavage site. Black boxes, identical amino acids; grey boxes, similar amino acids. Alignment created by GCG Pileup program, with modification. In the metalloprotease domain (red), amino acids that are critical for enzymatic activity are marked by asterisks. In the TSPt1 domain, amino acids that are conserved in vertebrate TSPt1 repeats are marked by plus symbols. Filled triangle, site of *gon-1(q518)*. Only the first two TSPt1-like motifs are shown; the consensus sequence for these repeats is W-X₄₋₅-W-X₂-CS-X₂-CG-X₄₋₅-X-G-X₃-R-X₃-C-X₄₋₂₇-C-X₈₋₁₂-C-X₃₋₄-C.

multiple thrombospondin type-1-like repeats. This motif architecture is typical of a small family of genes that include bovine procollagen I N-protease (PINP), which cleaves collagen³, and murine ADAMTS-1, the expression of which correlates with tumour cell progression⁴. We find that *gon-1* is expressed in two sites, leader cells and muscle, and that expression in each site has a unique role in forming the gonad. We speculate that GON-1 controls morphogenesis by remodelling basement membranes and that regulation of its activity is crucial for achieving organ shape.

The *C. elegans* hermaphrodite gonad has two U-shaped arms that extend from the somatic gonadal structures (Fig. 1a). The shape of the male gonad is similar except that it has one arm. During gonadogenesis, elongation of the growing gonadal arms is controlled by leader cells¹, which are called distal tip cells (DTCs) in hermaphrodites and the linker cell (LC) in males. The *gon-1* gene is required for both elongation of the gonadal arms and morphogenesis of somatic gonadal structures² (Fig. 1b). To investigate *gon-1* function, we cloned the gene by a combination of genetic mapping, mutant rescue and RNA-mediated interference (Fig. 2a, legend). We confirmed the predicted gene F25H8.3 as *gon-1* by identifying molecular lesions in five *gon-1* mutant alleles (Fig. 2b, c; see below). The GON-1 protein bears a signal sequence at its amino terminus, but no predicted transmembrane domain, suggesting that it is secreted (Fig. 2c).

The GON-1 protein possesses a predicted metalloprotease (MP) domain (Fig. 2c, red bar) which is most similar to members of the reprotolysin subfamily⁵. A potential furin cleavage site is located at the N-terminal border of the metalloprotease domain (Fig. 2c, green bar)^{6,7}. GON-1 and the reprotolysins share a common zinc-binding active site with the larger metzincin superfamily⁸. Amino-acid conservation within the active site, together with the known crystal structure of several superfamily members, reveals those amino acids that are essential for enzymatic activity (marked by asterisks in Fig. 2c)⁸. GON-1 has all amino acids implicated in catalysis and all but one implicated in structure of the active site. To test whether the GON-1 metalloprotease is essential for gene activity, we compared mutant rescue by one GON-1 transgene bearing the wild-type metalloprotease sequence to that of a second GON-1 transgene bearing a mutation at amino acid 425 within the active site. This mutation changes glutamic acid to alanine (E425A). In the metzincin family of metalloproteases, this mutation abolishes enzymatic activity without altering protein structure or stability^{9,10}. The two transgenes were introduced at the same concentration into *gon-1* mutants, and rescue of arm extension was scored. The wild-type metalloprotease was capable of rescue (5/6 lines), whereas the mutant metalloprotease was not (0/16 lines). We conclude that GON-1 is an active metalloprotease and that metalloprotease activity is essential for its role in controlling morphogenesis.

In addition to its metalloprotease domain, GON-1 possesses a series of thrombospondin type 1 (TSPT1) repeats (Fig. 2b, c). The most N-terminal TSPT1 repeat bears the hallmarks of a canonical vertebrate TSPT1 repeat (15/16 of the consensus amino acids; + in Fig. 2c)¹¹, whereas the remaining 17 repeats are less similar and define a TSPT1-like variant. The function of the TSPT1 repeats was explored by analysing *gon-1* nonsense mutations predicted to remove varying numbers of repeats (Fig. 2b). One mutation, *gon-1(q518)*, is in the canonical TSPT1 motif, and the others are in the TSPT1-like repeats. In *C. elegans*, mRNAs containing premature stop codons are normally degraded by the *smg* system, but, those mRNAs are stabilized in an *smg*-mutant background¹², the remaining activity of truncated GON-1 proteins should be evident in *smg-1*; *gon-1* double mutants. We found that *gon-1(q518)* was not suppressed in an *smg* background, whereas all four mutations in the TSPT1-like repeats (Fig. 2b) were suppressed (see Methods). Therefore, the GON-1(q518) mutant protein, which has the metalloprotease domain but not the bona fide TSPT1 motif or the rest of the C-terminal protein, is not capable of mutant rescue, but the other truncated

proteins are. We infer that the region spanning from the TSPT1 motif to the second TSPT1-like motif is critical for GON-1 activity. We confirmed that two TSPT1-like repeats are sufficient for rescuing activity by mutant rescue with a mini-transgene (see below).

To determine where and when GON-1 is expressed during development, we generated a transgene, called *gon-1 5'::GFP*, carrying a predicted *gon-1* promoter fused to the coding region for green fluorescent protein (GFP). *gon-1 5'::GFP* drives GFP expression in the leader cells of both hermaphrodites and males during the period of most active gonadal-arm elongation (Fig. 3). In hermaphrodites, GFP is not observed in gonads from first-stage (L1) larvae (Fig. 3a, b). Its first gonadal expression occurs in L2 and is limited to the DTCs (Fig. 3c, d). GFP continues to be expressed in DTCs through L4 (Fig. 3e, f), but is faint or not detectable in the adult (Fig. 3g, h). GFP is similarly expressed in the male linker cell (Fig. 3i, j; data not shown). In addition to its expression in leader cells, the *gon-1* promoter also drives GFP in muscle cells throughout development (Fig. 3k; data not shown).

To explore the significance of *gon-1* expression in leader cells and in muscle, we placed *gon-1* coding sequences under the control of either of two heterologous promoters. The *lag-2* promoter drives expression in the hermaphrodite DTC¹³, whereas the *unc-54* promoter drives expression in body-wall muscle¹⁴. For both transgenes, effects were examined in a *gon-1* null (0) background. We found that *lag-2 5'::GON-1* rescues arm extension and fertility (5/6 lines) (Fig. 4c), but *unc-54 5'::GON-1* does not (0/11 lines). However, *unc-54 5'::GON-1* does affect gonadal shape. Whereas the gonadal tissues of *gon-1* mutants are not consolidated in a single mass (Fig. 4b), those of *gon-1(0)*; *unc-54 5'::GON-1* form a discrete mass that

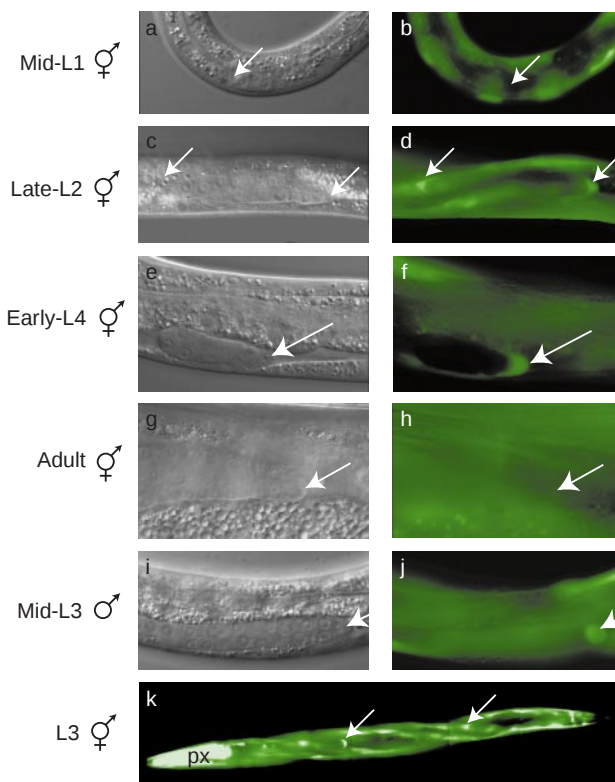


Figure 3 The *gon-1* promoter drives GFP expression in leader cells and in muscle. All animals are wild type and carry *gon-1 5'::GFP*. **a, c, e, f, h**, Nomarski; **b, d, f, g, i, j**, GFP fluorescence. Same animal in each left-right pair. Arrows, leader cells. **a, b**, No GFP in early gonad (arrow, progenitor of leader cell). **c-f**, GFP is expressed in leader cells during active elongation. **g, h**, GFP is not expressed in adult leader cell. **i, j**, Male leader cell expresses GFP. **k**, GFP is expressed in all muscles as well as in the leader cells. Body wall muscle extends length of body in four stripes, which are twisted because of the roller phenotype (see Methods). px, Pharynx; arrows, DTCs.

is expanded in all axes (Fig. 4d). Therefore, muscle expression of *gon-1* permits the contained growth and expansion of gonadal tissues within a basement membrane. Furthermore, because GON-1 expressed in muscle can act non-autonomously on the gonad, we conclude that GON-1 is secreted and can act over a distance. We propose that, in wild-type animals, *gon-1* must be expressed in leader cells to achieve a localized activity that directs elongation of gonadal arms along the proximal–distal axis. In addition, *gon-1* expression in muscle may provide a diffuse activity that is important for expansion of the gonad along dorsal–ventral and left–right axes. By this model, the combination of localized and dispersed activities is important in defining the final shape of the organ.

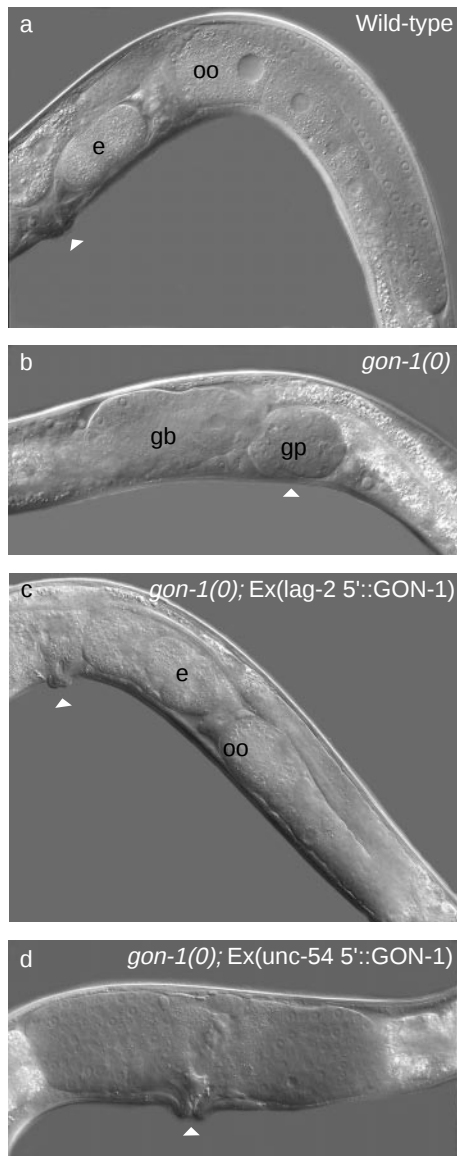


Figure 4 *gon-1* expression is critical for organ shape. **a–d**, Differential interference contrast (DIC) microscopy of adult hermaphrodites. Solid triangle, vulva; e, embryo; oo, oocyte. **a**, Wild type. Gonadal arm has normal U-shape. **b**, *gon-1(0)*. No arm extension and gonadal tissues are irregularly shaped. gp, gonad proper; gb, gonadal tissue bulging from gp. **c**, *gon-1(0)* with *gon-1(+)* expressed in DTC. Gonad can develop with U-shaped arm. Some animals are fertile, but more commonly, somatic gonadal structures (for example, uterus) are poorly formed and animals are sterile. Furthermore, the distal gonadal arms are thinner than normal. **d**, *gon-1(0)* with *gon-1(+)* expressed in body wall muscle. Gonadal tissue is more coherent than in *gon-1* mutants and has expanded in all three axes.

We suggest that the GON-1 metalloprotease permits and directs expansion of the gonad by remodelling the basement membrane. GON-1 belongs to a small gene family of secreted metalloproteases with one canonical TSPT1 domain, multiple TSPT1-like motifs and two conserved cysteine-rich regions (Fig. 2b). Based on this conserved architecture, we suggest the name MPT (for MetalloProtease with TSPT1 repeats) for the family. Bovine PINP can proteolyse the N-terminal propeptide from collagen I (refs 3, 15). We show that metalloprotease activity is required for GON-1 function and propose that, like PINP, it may cleave components of the extracellular matrix. Murine expression of *adamts-1* RNA correlates with tumour cell progression⁴. Given the role of GON-1 in promoting cell migration of the *C. elegans* leader cell, we suggest that MPT proteins may be involved more generally in cell migrations that must pass through extracellular matrix and that, in cancerous tissues, a change in MPT regulation may promote metastasis.

The GON-1 metalloprotease drives arm extension during gonadogenesis in *C. elegans*, and localization of its activity has a profound effect on organ shape. We suggest that similar activities may control organ morphogenesis throughout the animal kingdom, and previous *in vitro* experiments support this notion. For example, antibodies recognizing matrix metalloprotease 9 (MMP9) can block branching of the ureter bud during kidney development¹⁶, and inhibitors of MMPs block the invasion of endothelial cells into a fibrin matrix in assays for angiogenesis¹⁷. Based on these observations and our analysis of GON-1, we suggest that the matrix metalloproteases are critical modulators of organogenesis. With the identification of GON-1 in *C. elegans*, the functions of MPT metalloproteases can be analysed *in vivo* during organogenesis using the full force of molecular genetics available in this model system. □

Methods

Mapping and rescue of *gon-1*. Four-factor mapping placed *gon-1* between Tc1 insertions pKP614 and pKP5151. Specifically, two out of ten fertile Unc recombinants from *unc-24 gon-1 dpy-20/pKP5151* mothers did not carry the Tc1, placing *gon-1* right of pKP5151. One of 52 fertile Dpy recombinants from *unc-24 gon-1 dpy-20/pKP614* was negative for the Tc1, placing *gon-1* left of pKP614. Tc1 insertions were scored by polymerase chain reaction (PCR) as previously described¹⁸. For rescue experiments, 10 $\mu\text{g ml}^{-1}$ cosmid plus 100 $\mu\text{g ml}^{-1}$ pRF4 were co-injected into *unc-24 gon-1/gon-4 dpy-20* as described previously¹⁹. From 13 F1 roller lines injected with the T13H10 cosmid, we identified 3 rescued fertile Unc progeny. pJK651 (a deleted version of T13H10), was constructed by digesting with *KpnI* and religating the backbone. pJK651 rescued at a similar efficiency as T13H10. RNA-mediated interference (RNAi) was performed with both single-stranded and double-stranded RNA. Approximately 2 mg ml^{-1} RNA corresponding to a 1 kilobase (kb) portion of each predicted transcript on T13H10 were injected into N2 gonads as described²⁰. Phenocopy was scored as absence of DTC migration. Molecular lesions were identified by direct sequencing of PCR products spanning the *gon-1* gene from single-mutant worms²¹. Sequencing was done on an ABI sequencer. Complementary DNAs were isolated from the following libraries: λRB1 and λRB2 (gifts from R. Barstead) and $\lambda\text{AE.1}$ (gift from A. Puoti). cDNAs spanning the remaining gaps were isolated by performing PCR after reverse transcription of RNA²². The 5' end was identified using an SL1-specific primer and nested primers within the predicted transcript²². For *gon-1*; *smg-1* double mutants, non-Unc progeny from *gon-1/nT1[unc-(n754 dm) let-?]*; *smg-1(r861)* mothers were isolated and scored for fertility and arm extension. *nT1[unc-(n754 dm) let-?]* is a dominant Unc balancer for *gon-1*.

Transgene experiments. To construct *gon-1 5'::GFP*, we performed Expand High-Fidelity PCR (Boehringer Mannheim) with restriction-site-tagged primers that spanned –7 to –11,547 relative to the initiator ATG of the *gon-1* genomic DNA sequence. This region contains 4 kb of *gon-1* 5' flanking region and 8 kb of the *gon-1* transcribed region preceding the first AUG (two large introns, one small intron and the 500 nucleotides of 5'UTR (Fig. 2a)). All PCR reactions used pJK561 as template. The *gon-1 5'* PCR product was cloned into the pPD95.77 vector (from A. Fire). The resulting *gon-1 5'::GFP* construct was co-injected at 5 $\mu\text{g ml}^{-1}$ with pRF4 (100 $\mu\text{g ml}^{-1}$). Stably transmitting roller

lines were scored for fluorescence. To construct the *gon-1(+)* minigene, we performed high-fidelity PCR with restriction-site-tagged primers spanning -7 to +5,276 of *gon-1* gDNA. To construct the *gon-1* (E → A) mutant minigene, we made primers that introduced a T-to-G mutation at +2,546 and an A-to-C mutation at +2,551. The first mutation is translationally silent and introduces a *SphI* restriction site whereas the second mutation causes an E-to-A transition in the GON-1 metalloprotease active site. We used these primers to amplify either half of the wild-type minigene and spliced the resulting products using the new *SphI* site. Sequencing of *gon-1(+)* and *gon-1* (E → A) showed no second-site mutations. The two minigenes were cloned into pPD49.26 (from A. Fire) together with the *lag-2* 5' promoter from pJK375 (ref. 13). The *lag-2* 5':GON-1(+) and *lag-2* 5':GON-1(E → A) constructs were injected at 1 $\mu\text{g ml}^{-1}$ together with pRF4 (50 $\mu\text{g ml}^{-1}$) into *unc-24(e138) gon-1(q518)/gon-4(e2575) dpy-20(e1282)* hermaphrodites. Roller Unc animals from stably transmitting roller lines were scored for arm extension and fertility in both F2 and F3 generations. The *unc-54::GON-1(+)* transgene was constructed by cutting the *gon-1(+)* minigene from *lag-2* 5':GON-1(+) and splicing it downstream of the *unc-54* promoter in pPD30.35 (from A. Fire). The resulting transgene was co-injected at both 1 $\mu\text{g ml}^{-1}$ (50 $\mu\text{g ml}^{-1}$ pRF4) and 5 $\mu\text{g ml}^{-1}$ (100 $\mu\text{g ml}^{-1}$ pRF4) into *gon-1* heterozygotes and scored as above. Effects on *gon-1* gonadal structure were mild at 1 $\mu\text{g ml}^{-1}$ and stronger at 5 $\mu\text{g ml}^{-1}$.

Received 24 March; accepted 27 April 1999.

- Kimble, J. E. & White, J. G. On the control of germ cell development in *Caenorhabditis elegans*. *Dev. Biol.* **81**, 208–219 (1981).
- Belloch, R., Santa Anna-Arriola, S., Li, Y., Hodgkin, J. & Kimble, J. The *gon-1* gene regulates morphogenesis of the *C. elegans* gonad (submitted).
- Colige, A. *et al.* cDNA cloning and expression of bovine procollagen I N-proteinase: a new member of the superfamily of zinc-metalloproteinases with binding sites for cells and other matrix components. *Proc. Natl Acad. Sci. USA* **94**, 2374–2379 (1997).
- Kuno, K. *et al.* Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene. *J. Biol. Chem.* **272**, 556–562 (1997).
- Rawlings, N. D. & Barrett, A. J. Evolutionary families of metalloproteinases. *Methods Enzymol.* **248**, 183–228 (1995).
- Pei, D. & Weiss, S. J. Furin-dependent intracellular activation of the human stromelysin-3 zymogen. *Nature* **375**, 244–247 (1995).
- Pei, D. & Weiss, S. J. Transmembrane-deletion mutants of the membrane-type matrix metalloproteinase-1 process progelatinase A and express intrinsic matrix-degrading activity. *J. Biol. Chem.* **271**, 9135–9140 (1996).
- Stöcker, W. *et al.* The metzincins—topological and sequential relations between the astacins, adamalysins, serrins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.* **4**, 823–840 (1995).
- Cha, J. & Auld, D. S. Site-directed mutagenesis of the active site glutamate in human matrilysin: investigation of its role in catalysis. *Biochemistry* **36**, 16019–16024 (1997).
- Will, H., Atkinson, S. J., Butler, G. S., Smith, B. & Murphy, G. The soluble catalytic domain of membrane type 1 matrix metalloproteinase cleaves the propeptide of progelatinase A and initiates autolytic activation. *J. Biol. Chem.* **271**, 17119–17123 (1996).
- Adams, J. C., Tucker, P. R. & Lawler, J. *The Thrombospondin Gene Family* (Landes, Austin, Texas, 1995).
- Anderson, P. & Kimble, J. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 185–208 (Cold Spring Harbor Laboratory Press, 1997).
- Henderson, S. T., Gao, D., Lambie, E. J. & Kimble, J. *lag-2* may encode a signaling ligand for the GLP-1 and LIN-12 receptors of *C. elegans*. *Development* **120**, 2913–2924 (1994).
- Okkema, P. G., Harrison, S. W., Plunger, V., Aryana, A. & Fire, A. Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics* **135**, 385–404 (1993).
- Colige, A. *et al.* Characterization and partial amino acid sequencing of a 107-kDa procollagen I N-proteinase purified by affinity chromatography on immobilized type XIV collagen. *J. Biol. Chem.* **270**, 16724–16730 (1995).
- Lelongt, B., Trugnan, G., Murphy, G. & Ronco, P. M. Matrix metalloproteinases MMP2 and MMP9 are produced in early stages of kidney morphogenesis but only MMP9 is required for renal organogenesis in vitro. *J. Cell Biol.* **136**, 1363–1373 (1997).
- Hiraoka, N., Allen, E., Apel, I. J., Gyetko, M. R. & Weiss, S. J. Matrix metalloproteinases regulate neovascularization by acting as pericellular fibrinolysins. *Cell* **95**, 365–377 (1998).
- Korswagen, H. C., Durbin, R. M., Smits, M. T. & lasterk, R. H. A. Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. *Proc. Natl Acad. Sci. USA* **93**, 14680–14685 (1996).
- Mello, C. & Fire, A. in *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 451–482 (Academic, New York, 1995).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Williams, B. D., Schrank, B., Huynh, C., Showkneen, R. & Waterston, R. H. A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics* **131**, 609–624 (1992).
- Krause, M. in *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 513–533 (Academic, New York, 1995).
- Schedl, T. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 241–269 (Cold Spring Harbor Laboratory Press, 1997).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).

Acknowledgements. We thank A. Coulson for providing cosmid, R. Barstead and A. Puoti for cDNA libraries, A. Fire for GFP and expression vectors, D. Gao for *lag-2* promoter, J. Nance for Tc1 primers, the *Caenorhabditis* Genetics Center for worm strains, and the many lab members who provided thoughts and criticisms during the course of this work. R.B. is an MD/PhD student and was an NIH Molecular Biosciences trainee. J.K. is an investigator with the Howard Hughes Medical Institute and has been supported by grants from NIH and NSF.

Correspondence and requests for materials should be addressed to J.K. (e-mail: jekimble@facstaff.wisc.edu).

Emergence of vancomycin tolerance in *Streptococcus pneumoniae*

R. Novak*, B. Henriques†, E. Charpentier*, S. Normark† & E. Tuomanen*

* Dept of Infectious Diseases, St. Jude Children's Research Hospital, 332 N. Lauderdale St., Memphis, Tennessee 38105, USA

† Swedish Institute for Infectious Disease Control, Karolinska Institute, S171-82 Solna, Stockholm, Sweden

Streptococcus pneumoniae, the pneumococcus, is the most common cause of sepsis and meningitis¹. Multiple-antibiotic-resistant strains are widespread, and vancomycin is the antibiotic of last resort^{2,3}. Emergence of vancomycin resistance in this community-acquired bacterium would be catastrophic. Antibiotic tolerance, the ability of bacteria to survive but not grow in the presence of antibiotics, is a precursor phenotype to resistance⁴. Here we show that loss of function of the VncS histidine kinase of a two-component sensor-regulator system in *S. pneumoniae* produced tolerance to vancomycin and other classes of antibiotic. Bacterial two-component systems monitor environmental parameters through a sensor histidine-kinase/phosphatase, which phosphorylates/dephosphorylates a response regulator that in turn mediates changes in gene expression. These results indicate that signal transduction is critical for the bactericidal activity of antibiotics. Experimental meningitis caused by the *vncS* mutant failed to respond to vancomycin. Clinical isolates tolerant to vancomycin were identified and DNA sequencing revealed nucleotide alterations in *vncS*. We conclude that broad antibiotic tolerance of *S. pneumoniae* has emerged in the community by a molecular mechanism that eliminates sensitivity to the current cornerstone of therapy, vancomycin.

The bactericidal activity of antibiotics that stop cell-wall synthesis relies on activation of bacterially encoded death effectors^{4–6}. In pneumococcus, the autolysin protein LytA serves this function and is triggered by antibiotics to digest the cell-wall exoskeleton⁴. The autolysins of all bacteria, including the pneumococcal LytA, are presumed to be under strong negative regulation as they are constitutively produced, potentially suicidal and physiologically activated in stationary phase⁶. The mechanism of their activation is unknown and is an important question in basic microbiology. To define elements in the autolysin trigger pathway, we searched for loss of penicillin-induced autolysin in a library of pneumococcal

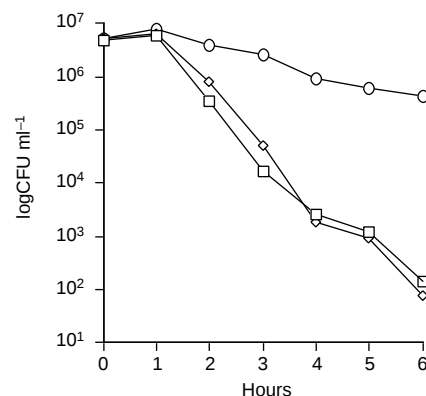


Figure 1 Effect of loss of function of *vncR* and *vncS* on bactericidal activity of vancomycin. Wild-type strain R6 (squares), *vncR* mutant (diamonds) and *vncS* mutant (circles). Cultures in the early exponential phase of growth (10⁷ CFU ml⁻¹) were treated with 10 × the MIC of vancomycin (5 $\mu\text{g ml}^{-1}$), and bacterial viability was followed for 6 h.

mutants with loss-of-function mutations saturating the chromosome^{7,8}. The screen identified 17 independent mutants that contained active autolysin but failed to die in the presence of penicillin. This property is termed tolerance⁴. Mutant SPSJ01 was also tolerant to vancomycin (Fig. 1). SPSJ01 had an interruption in a putative histidine kinase of a two-component regulatory system.

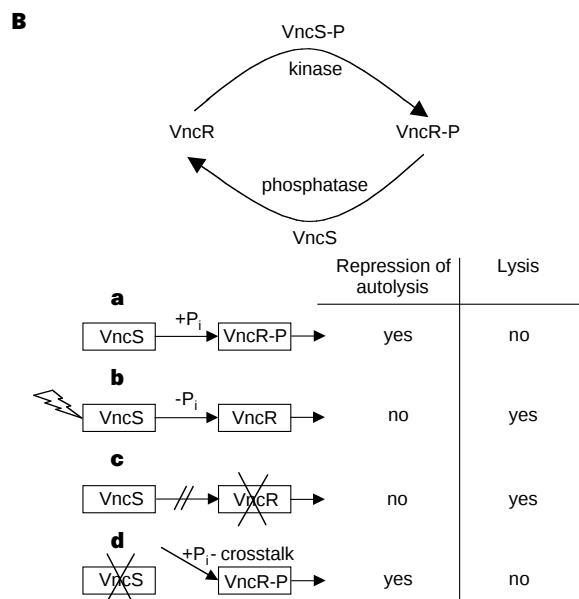
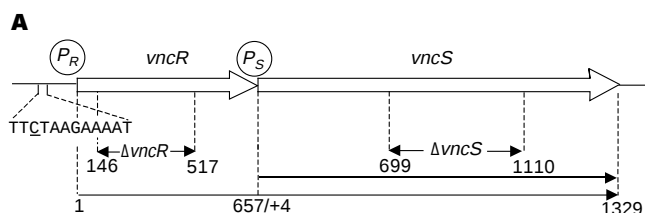


Figure 2 Interrelationships of VncR and VncS. **A**, Organization of the two-component regulatory system, VncS–VncR. Putative promoters (P_R and P_S) are at positions –18 and 642. A 12-bp DNA-binding site for phosphorylated VanR of *E. faecium*²⁶ was present 35-bp upstream of the putative promoter region of *vncR*. Northern-blot analysis of the parent strain R6 total RNA probed with an intragenic fragment of *vncR* revealed a transcript of 2 kb indicating readthrough of *vncS*. A probe specific for *vncS* revealed a transcript of 1.3 kb matching the size of *vncS*. The C-terminal portion of VncS contained four of five blocks of conserved amino acids characteristic for transmitter modules of histidine kinases²⁷. The D-box, which is likely to be involved in binding of Mg, was missing, a finding similar to VanS_B of *E. faecium*²⁶. The predicted site for autophosphorylation was found at amino-acid positions 239–246, X₂HEX_{3–4}P, and was identical to the site in VanS_B. As the first 170 amino acids of VncS contain two clusters of hydrophobic amino acids, which could correspond to membrane-spanning regions, VncS is probably a membrane protein. On the basis of sequence similarity, the VncR protein can be assigned to the OmpR subclass of response regulators^{28,29}. Alignment of the DNA-binding domain of VncS and VanR, which also belongs to the OmpR family, demonstrated a nearly identical match of the residues corresponding to the hydrophobic core of OmpRc. **B**, Model of the VncS/VncR signal transduction pathway. The histidine kinase/phosphatase VncS controls the level of phosphorylation of the response regulator VncR. **a**, VncR-P represses the activation of autolysin. **b**, A stimulus sensed by VncS activates its phosphatase activity and dephosphorylates VncR-P, and VncR enables a pathway of autolysis, for example during stationary phase or following triggering by antibiotics. **c**, Genetic loss of VncR simulates lack of the repressor VncR-P; therefore the autolytic pathway can be activated. **d**, Genetic loss of VncS eliminates phosphatase activity, allowing continued repression of autolysis by VncR-P.

Inverse polymerase chain reactions and sequence analysis revealed an operon encoding the histidine kinase, named *vncS*, and a contiguous response regulator, *vncR* (Fig. 2A). To investigate the role of VncS–VncR in promoting vancomycin tolerance, we used insertion-duplication mutagenesis to create loss of function in either of the *vncS* and *vncR* genes (Fig. 2A). Upon exposure to 10 times the minimum inhibitory concentration (MIC: 5 $\mu\text{g ml}^{-1}$) of vancomycin, the parent strain R6 and the *vncR* mutant underwent autolysis (4 log loss in viability over 4 h). In contrast, the *vncS* mutant stopped growing but showed virtually no loss in viability (Fig. 1). Penicillin and vancomycin retained access to their respective targets, as indicated by unchanged MICs of 0.01 and 0.5 $\mu\text{g ml}^{-1}$, respectively. Tolerance of the *vncS* mutant extended to β -lactam antibiotics, cephalosporins, aminoglycosides and quinolones (data not shown), indicating that signal transduction through VncS is required for bacterial death in response to a broad range of mechanistically distinct antibiotics.

The deduced amino-acid sequences of VncS and VncR showed 38% identity to the VanS_B–VanR_B two-component regulatory system encoded on plasmids or large conjugative chromosomal elements of glycopeptide-resistant *Enterococcus faecalis*⁹. VanR_B activates transcription of the *vanH_B*, *vanB* and *vanX_B* genes, resulting in incorporation of D-alanyl-D-lactate instead of D-alanyl-D-alanine into the peptidoglycan. This change leads to a reduction in the affinity of vancomycin for the cell-wall target¹⁰. Scanning of the pneumococcal genome revealed no homologues of *E. faecalis* *vanH_B*, *vanB* or *vanX_B*. Thus, although similar two-component systems affect susceptibility to vancomycin in two Gram-positive bacteria, their underlying roles in resistance and tolerance appear to be distinct, presumably reflecting different gene products controlled by the response regulator. Although VanB-type vancomycin resistance is relatively rare in enterococci and resistance arises from gain of function of the *vanB* gene cluster, all pneumococci appear to harbour chromosomal *vncS*–*vncR* (data not shown), and tolerance results from loss of function of VncS. However, the mechanism of signal transduction may be similar in VanS_B/VanR_B and VncS/VncR. Rather than the traditional histidine kinase, VanS_B acts as a phosphatase, deactivating VanR_B–P¹¹. We propose that VncS may also dephosphorylate VncR and that autolysin activity is repressed

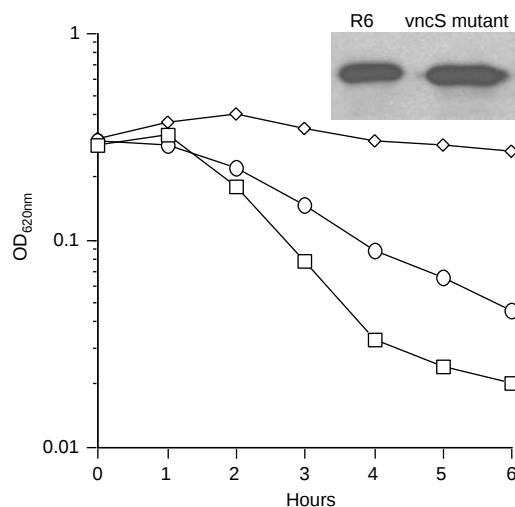


Figure 3 Functional assay of autolytic activity. Autolysin preparations^{4,7} of parent strain R6 (squares) (positive control), *vncS* mutant (circles) and the autolysin-defective strain Lyt-4-4 (diamonds) (negative control) were added to cultures of Lyt-4-4 at an OD_{620nm} of 0.3. Lysis was followed after addition of 10 $\times$ MIC of vancomycin (5 $\mu\text{g ml}^{-1}$) (data were shown as the mean out of three independent experiments, s.d. $\pm$ 10%). Inset, Immunoblot analysis of autolysin preparations of the *vncS* mutant. Lysates of the *vncS* mutant were transferred to nitrocellulose and probed with rabbit anti-autolysin antibody (1:1000). Parent strain R6 was used as a positive control.

by phosphorylated VncR (Fig. 2B). Relief of repression occurs upon a stimulus which activates the phosphatase activity of VncS (Fig. 2B). This is consistent with findings that pneumococci lacking VncS are tolerant (cannot relieve VncR–P repression), but mutants lacking VncR are not (absence of VncR–P allows lysis).

The original description of tolerance identified the critical role of the autolysin LytA in the bactericidal activity of penicillin against pneumococci⁴. A simple explanation for tolerance of the *vncS* mutant would be loss of the expression or intrinsic activity of LytA. However, western-blot analysis of LytA in the *vncS* mutant showed normal amounts of secreted protein (Fig. 3). LytA activity was also intact, as shown by the ability of autolysin isolated from both wild-type and VncS-deficient pneumococci to reconstitute lysis of the autolysin-deficient strain Lyt-4-4 (Fig. 3). The lack of autolysis of the *vncS* mutant upon addition of vancomycin is thus due to an alteration in the control of the autolytic pathway, rather than in the amount or inherent activity of autolysin itself.

Pneumococcus is naturally transformable¹² and this property is critical to the evolution of antibiotic resistance^{13,14}. It has been suggested that tolerance may favour the development of drug resistance, because it creates survivors of antibiotic therapy¹⁵. Comparison between the parent R6 and the *vncS* mutant for transformation of DNA encoding streptomycin or penicillin resistance consistently indicated greater efficiency for the *vncS* mutant. Transformation to streptomycin resistance occurs in one round of transformation, the efficiency of which was doubled in the *vncS* mutant (0.2% for R6 and 0.41% for *vncS* mutant). Transformation to high-level penicillin resistance requires several rounds of transformation^{13,16}, some affecting the penicillin-binding proteins and some altering unknown accessory genes¹⁷, potentially tolerance genes. The parent R6 was not able to acquire the ability to survive an increase in penicillin concentration from 6 to 50 ng ml⁻¹ in one round of transformation. In contrast, the *vncS* mutant consistently achieved this increase in penicillin resistance in one transformation step (0 versus 0.005% for *vncS* mutant). These data are consistent with the hypothesis that tolerance may be a preferred background for the import of resistance mutations.

Morbidity (>50%) and mortality (30%) due to pneumococcal meningitis are greater than for all other meningeal pathogens¹. Because of the lack of host defences in the cerebrospinal fluid, such infections must be treated with bactericidal antibiotics¹⁸. Reduced killing properties, as in the *vncS* mutant, would theoretically lead to persistence of the infectious agent. To test this, meningitis due to the vancomycin-tolerant *vncS* mutant was induced in the established rabbit meningitis model¹⁹. Rabbits were inoculated intrathecally with 10⁵ bacteria using (1) a vancomycin-sensitive encapsulated pneumococcus, D39; (2) an isogenic vancomycin-tolerant strain, D39Δ*vncS*; or (3) an equal mixture of the two. Twenty-four hours later, untreated rabbits in each group had acute bacterial meningitis. This was also the case when the strains were co-injected to cause a mixed infection, indicating equivalent growth potential (Table 1). Animals were treated intravenously with vancomycin (30 mg kg⁻¹) at 6 and 12 h. Analysis of cerebrospinal fluid showed that the parent strain was readily killed (Table 1). In contrast, the density of the tolerant strain D39Δ*vncS* remained constant at 10⁵ bacteria per ml cerebrospinal fluid even after the second vancomycin bolus (Table 1). Twelve hours after the last vancomycin injection, the number of tolerant bacteria increased to 10⁷ bacteria per ml, indicating that therapy had failed.

The animal studies indicated that the existence of clinical isolates tolerant to antibiotics might lead to complications of therapy, particularly in meningitis. To screen for the existence of such strains in the community, 116 clinical isolates were tested for killing in the presence of 10 × the minimum inhibitory concentration (MIC) of vancomycin. Vancomycin tolerance was defined as killing of ≤ 2 ± 0.6 log bacteria in 4 h, a range more than 1 s.d. from the response of the wild type and within 1 s.d. of the standard tolerant

Table 1 Effect of vancomycin treatment on bacterial density in CSF

		Mean bacterial density in CSF × 1,000		
Bacteria*	Vancomycin†	6 h	12 h	24 h
Single strain infection				
D39	–	28.5 ± 8.5	600 ± 200	8250 ± 650
D39	+	44.5 ± 5.5	0.2 ± 0.02	0 ± 0
D39Δ <i>vncS</i>	–	55.0 ± 15.0	495 ± 105	7945 ± 1055
D39Δ <i>vncS</i>	+	30.5 ± 7.5	18 ± 6.0	1000 ± 200
Mixed infection (D39 + D39Δ <i>vncS</i>)				
D39	+	23.0 ± 8	0.5 ± 0.1	0 ± 0
D39Δ <i>vncS</i>	+	27.5 ± 15.5	6.85 ± 0.15	1700 ± 600

*n, four rabbits per group
†30 mg kg⁻¹ i.v. at 6 and 12 h

strain Lyt-4-4 (ref. 20). Vancomycin tolerance was found in three pneumococcal strains, all belonging to serotype 9V. The three isolates were resistant to penicillin. Sequence analysis of *vncS* revealed a substitution of valine for alanine at position 440 in all three tolerant strains compared to six non-tolerant strains. Complementation of the *vncS* variant in the tolerant clinical isolates with a wild-type allele restored lysis. Similarly, complementation of a knockout *vncS* mutant with the wild-type *vncS* gene restored lysis, linking the regulatory pathway governed by VncS to the tolerant phenotype in laboratory and clinical isolates. Thus, the changes in VncS identified in the clinical isolates appear to at least mark tolerant strains. This does not rule out that tolerance in the variety of backgrounds seen in clinical isolates is multigenic²¹.

In conclusion, vancomycin tolerance has emerged in *S. pneumoniae* and tolerant strains may be more easily transformed to high-level antibiotic resistance. A molecular mechanism for tolerance involves the functional loss of VncS, a histidine kinase/phosphatase related in its sequence but not in its pathway to that found on plasmids or on large conjugative chromosomal elements in type B vancomycin-resistant *E. faecalis*. This pneumococcal two-component system may regulate a basic pathway triggering autolysis (Fig. 2B), a pathway that is physiologically activated only in stationary phase but is unnaturally activated by antibiotics. Our data indicate that signal transduction is necessary for antibiotic-induced killing and that there is a multi-component triggering pathway. The data obtained from the rabbit meningitis model indicate that, if tolerant strains invade the central nervous system, they may be responsible for relapsing or persistent bacterial meningitis. Currently, clinical laboratories cannot detect antibiotic-tolerant strains without performing cumbersome kill curves²². The correlation of the changes in sequence of *vncS* and clinical tolerance suggests that probes for *vncS* may serve as markers for rapidly tracking the tolerance trait. □

Methods

Bacterial strains and growth conditions. The parental strain of *S. pneumoniae* D39, serotype 2 and its unencapsulated derivatives R6 and Lyt-4-4 were grown on tryptic soy agar (TSA, Difco) supplemented with sheep blood to a final concentration of 3% (v/v)⁴. For growth in liquid culture, the bacteria were grown in a semi-synthetic casein hydrolysate medium supplemented with yeast extract²³. For maintenance of the chromosomally integrated plasmid pJDC9::*vncS/vncR*, bacteria were grown in the presence of 1 μg/ml erythromycin (Sigma).

Recombinant DNA methods, insertion duplication mutagenesis of D39 and complementation. DNA ligations, restriction endonuclease digestions, agarose gel electrophoresis and DNA amplification by PCR were performed according to standard techniques²⁴. DNA purification and plasmid preparations used kits from Qiagen and Promega/Wizard according to the manufacturer's instructions. Transformation of *E. coli* with plasmid DNA was carried out with CaCl₂-treated cells as described²⁵. Transformation of *S. pneumoniae* was performed according to standard protocols⁷. For insertional duplication mutagenesis of *vncS*, an internal 411-bp fragment (bp 699–1110) was amplified using total DNA of D39 and the primers eVNCs (5'-AAG TTG ATT TCA AGG

ACA GTC TTC CT-3') and bVNCs (5'-AGA ATG GAT CCA TTC TTT TCA TGC AA-3'). Insertional duplication of *vncR* was done by amplifying a 371-bp fragment (bp 146–517) using the primers eVNCr (5'-TAC TGG AAT TCC AGA TGC CCA AGC TC-3') and bVNCr (5'-ATC TGG GAT CCA GTC AAG GCC CGG CC-3'). The PCR products were digested with *EcoRI* and *BamHI*. We cloned the amplified fragments in pJDC9 (ref. 25). The resulting recombinant plasmids were then transformed into D39 and R6, respectively. For complementation experiments, a DNA fragment containing the wild type *vncS* gene placed downstream of the *comA* inducible promoter was cloned into the shuttle vector pMU1328. We used the resulting construct to transform the tolerant clinical isolate and the *vncS* knockout mutant. *comA* was amplified using the primers KpnIComA (5'-TAT TGA GGT ACC TTT GGC GAT GAT TTC C-3') and NdeIComA (5'-AAA TTT CAT ATG CAT CTC CTT TTC CC-3'). The wild-type and mutated *vncS* gene were amplified using the primers BamHIVncS (5'-TCT TCT GGA TCC ATC TAT GTA AAC CC-3') and NdeIVncS (5'-TGG AGC GCATAT GAA ACG AAC AGG-3'). DNA sequencing was done at the Center for Biotechnology at St. Jude Hospital. The SSCP (single-strand conformation polymorphism) analysis was performed using five pairs of primers generating 200-bp products across the *vncS* gene according to the manufacturer's protocol (FMC).

DNA transformation. Comparison of the efficiency of transformation was done as described²¹ using 0.2 µg ml⁻¹ chromosomal DNA encoding streptomycin resistance or penicillin resistance from the highly resistant clinical strain 8249 (MIC 6 µg ml⁻¹)¹⁴. Selection for penicillin resistance above the MIC of 0.006 µg ml⁻¹ of the parent strain R6 was made at 0.012, 0.025 and 0.05 µg ml⁻¹ penicillin, values previously shown to be the minimal increments possible in a penicillin-sensitive background.

Vancomycin susceptibility and autolysis rates. We determined autolysis rates and viability using 10 ml cultures of *S. pneumoniae* exposed to 10 × MIC of vancomycin (5 µg ml⁻¹) when the OD_{620nm} reached 0.25 to 0.3 (corresponding to 5 × 10⁷ CFU ml⁻¹). After exposure times, 100 µl aliquots were serially diluted in semisynthetic medium and plated for viable counts on tryptic-soy agar containing 3% sheep blood (v/v).

Rabbit model of pneumococcal meningitis. Pneumococcal meningitis was established according to a standard protocol in three groups of four rabbits¹⁷. After withdrawal of CSF (300 µl), 10⁵ bacteria in 200 µl saline was introduced intracisternally: D39, D39Δ*vncS*, and D39 and D39Δ*vncS* simultaneously. Half of the rabbits in groups one and two received vancomycin (30 mg kg⁻¹) 6 and 12 h after infection; all rabbits in group three received vancomycin after 6 and 12 h. At 6 h intervals, CSF (200 µl) and blood were sampled and tested for bacterial and leukocyte density.

Received 22 March; accepted 9 April 1999.

- Schuchat, A. *et al.* Bacterial meningitis in the United States in 1995. Active Surveillance Team. *N. Engl. J. Med.* **337**, 970–976 (1997).
- Friedland, I. R. & Istre, G. R. Management of penicillin-resistant pneumococcal infections. *Pediatr. Infect. Dis. J.* **11**, 433–435 (1992).
- American Academy of Pediatrics in *Red Book: Report of the Committee on Infectious Diseases* 24th edn (ed. Peter, G.) 410–419 (Elk Grove Village, 1997).
- Tomasz, A., Albino, A. & Zanati, E. Multiple antibiotic resistance in a bacterium with suppressed autolytic system. *Nature* **227**, 138–140 (1970).
- Moreillon, P., Markiewicz, Z., Nachman, S. & Tomasz, A. Two bactericidal targets for penicillin in pneumococci: autolysis-dependent and autolysis-independent killing mechanisms. *Antimicrob. Agents Chemother.* **34**, 33–39 (1990).
- Holtje, J. V. & Tuomanen, E. I. The murein hydrolases of *Escherichia coli*: properties, functions and impact on the course of infections in vivo. *J. Gen. Microbiol.* **137**, 441–454 (1991).
- Pearce, B. J., Yin, Y. B. & Masure, H. R. Genetic identification of exported proteins in *Streptococcus pneumoniae*. *Mol. Microbiol.* **9**, 1037–1050 (1993).
- Williamson, R. & Tomasz, A. Antibiotic-tolerant mutants of *Streptococcus pneumoniae* that are not deficient in autolytic activity. *J. Bacteriol.* **144**, 105–113 (1980).
- Evers, S. & Courvalin, P. Regulation of VanB-type vancomycin resistance gene expression by the VanS_B two-component regulatory system in *Enterococcus faecalis* V583. *J. Bacteriol.* **178**, 1302–1309 (1996).
- Bugg, T. D. H. *et al.* Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry* **30**, 10408–10415 (1991).
- Silva, J. C., Haldimann, A., Prahalad, M. K., Walsh, C. T. & Wanner, B. L. In vivo characterization of the type A and B vancomycin-resistant enterococci (VRE) VanRS two-component systems in *Escherichia coli*: a nonpathogenic model for studying the VRE signal transduction pathways. *Proc. Natl Acad. Sci. USA* **95**, 11951–11956 (1998).
- Tomasz, A. & Hotchkiss, R. D. Regulation of the transformability of pneumococcal cultures by macromolecular cell products. *Proc. Natl Acad. Sci. USA* **51**, 480–487 (1964).
- Shockey, T. E. & Hotchkiss, R. D. Stepwise introduction of transformable penicillin resistance in pneumococcus. *Genetics* **64**, 397–408 (1970).
- Dowson, C. G. *et al.* Horizontal gene transfer and the evolution of resistance and virulence determinants in *Streptococcus*. *Soc. Appl. Bacteriol. Symp. Ser.* **26**, 425–515 (1997).
- Liu, H. H. & Tomasz, A. Penicillin tolerance in multiply drug-resistant natural isolates of *Streptococcus pneumoniae*. *J. Infect. Dis.* **152**, 365–372 (1985).

- Zigheboim, S. & Tomasz, A. Multiple antibiotic resistance in South African strains of *Streptococcus pneumoniae*: Mechanism of resistance to beta-lactam antibiotics. *Rev. Inf. Dis.* **3**, 267–276 (1981).
- Murakami, K. & Tomasz, A. Involvement of multiple genetic determinants in high-level methicillin resistance in *Staphylococcus aureus*. *J. Bacteriol.* **171**, 874–879 (1989).
- Quagliarello, V. & Scheld, W. Bacterial meningitis: pathogenesis, pathophysiology and progress. *N. Engl. J. Med.* **327**, 864–872 (1992).
- Tauber, M. G. & Zwaalen, A. Animal models of meningitis. *Methods Enzymol.* **235**, 92–106 (1994).
- Tuomanen, E. *et al.* Microbiological and clinical significance of a new property of defective lysis in clinical strains of pneumococci. *J. Infect. Dis.* **158**, 36–43 (1988).
- Novak, R., Braun, J. S., Charpentier, E. & Tuomanen, E. Penicillin tolerance genes of *Streptococcus pneumoniae*: The ABC-type manganese complex Psa. *Mol. Microbiol.* **29**, 1285–1296 (1998).
- CDC: Interim guidelines for prevention and control of staphylococcal infection associated with reduced susceptibility to vancomycin. *Mor. Mortal. Weekly Rep.* **46**, 626–635 (1997).
- Lacks, S. & Hotchkiss, R. D. A study of the genetic material determining an enzyme activity in pneumococcus. *Biochim. Biophys. Acta* **39**, 508–517 (1960).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning. A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1990).
- Chen, J. D. & Morrison, D. A. Cloning of *Streptococcus pneumoniae* DNA fragments in *Escherichia coli* requires vectors protected by strong transcriptional terminators. *Gene* **55**, 179–187 (1987).
- Holman, T. R., Wu, Z., Wanner, B. L. & Walsh, C. T. Identification of the DNA-binding site for the phosphorylated VanR protein required for vancomycin resistance in *Enterococcus faecium*. *Biochemistry* **33**, 4625–4631 (1994).
- Parkinson, J. S. & Kofoid, E. C. Communication modules in bacterial signaling proteins. *Annu. Rev. Genet.* **26**, 71–112 (1992).
- Stock, J. B., Ninfa, A. J. & Stock, A. M. Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol. Rev.* **53**, 450–490 (1989).
- Martinez-Hackert, E. & Stock, A. M. The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor. *Structure* **5**, 109–124 (1997).

Acknowledgements. This work was supported in part by grants from the NIH (to E.T.), the Cancer Center CORE Grant and the American Lebanese Syrian Associated Charities and Åke Wiberg Foundation. We thank J. S. Braun for help with the animal model, J. Li for technical assistance and P. Murray for helpful suggestions and discussions.

Correspondence and requests should be addressed to E.T. (e-mail: elaine.tuomanen@stjude.org). The *vncS* and *vncR* sequences are deposited in GenBank under accession number AF140356.

Modelling T-cell memory by genetic marking of memory T cells *in vivo*

Joshy Jacob* & David Baltimore*

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Immunological memory is the ability of the immune system to respond with enhanced vigour to pathogens that have been encountered in the past. Following infection or immunization, most effector T cells undergo apoptotic cell death, but a small fraction of these cells, proportional to the early antigen load and initial clonal burst size¹, persist in the host as a stable pool of memory T cells^{2–7}. The existence of immunological memory has been recognized for over 2,000 years, but our understanding of this phenomenon is limited, primarily because memory lymphocytes cannot be unequivocally identified as they lack specific, permanent markers. Here we have developed a transgenic mouse model system whereby memory T cells and their precursors can be irreversibly marked with a reporter gene and thus can be unambiguously identified. Adoptive transfer of marked CD8⁺ T cells specific for lymphocytic choriomeningitis virus protected naive recipients following viral challenge, demonstrating that we have marked memory T cells. We also show that cytotoxic effector lymphocytes that develop into memory T cells can be identified in the primary response.

In our transgenic mouse model system, a Cre-recombinase-mediated genetic recombination⁸ event is activated in antigen-stimulated T cells, leading to irreversible surface expression of the gene product of the reporter gene, human placental alkaline phosphatase (PLAP)⁹. Briefly, the strategy involves generating two independent transgenic mouse lines (Fig. 1): one has a truncated human granzyme-B promoter¹⁰ that drives expression of the Cre-recombinase enzyme specifically in activated T cells (granzyme-B–Cre;

* Present addresses: Emory Vaccine Center and the Department of Microbiology and Immunology, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA (J.J.); and California Institute of Technology, Pasadena, California 91125, USA (D.B.).

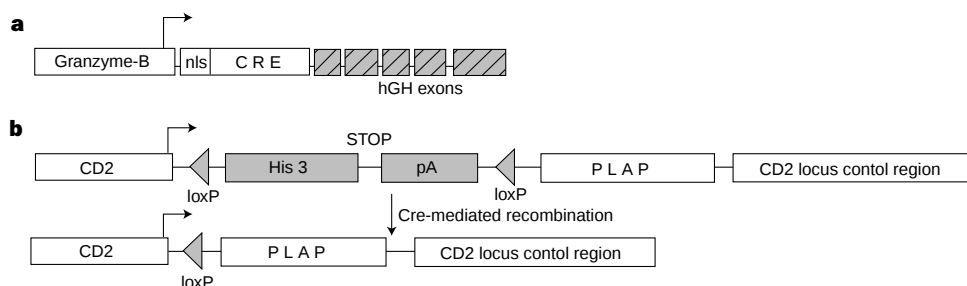


Figure 1 Transgenic constructs used. **a**, Granzyme-B-promoter-driven Cre recombinase (granzyme-B-Cre). The Cre recombinase contains a nuclear localization signal sequence (nls) of the SV40 large T antigen. The human growth hormone (hGH) polyadenylation sequence provides the intron-exon structure for optimal transgene expression. **b**, CD2-STOP-PLAP. The STOP

fragment contains the yeast His3 gene with polyadenylation sequence, splice donor and a false translational initiation sequence. *loxP* sites denote the 34-base pair (bp) Cre recombinase recognition sequence. Following Cre-mediated excisional recombination, PLAP is positioned adjacent to the CD2 promoter which is active at all stages of T-cell development.

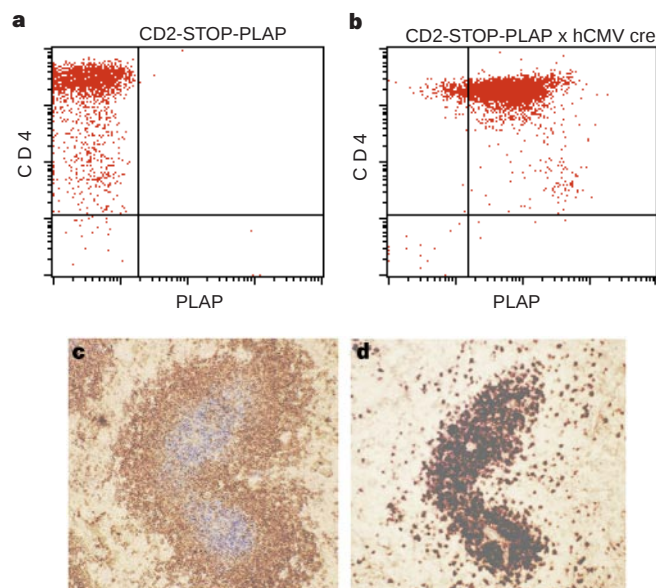


Figure 2 PLAP expression occurs only after Cre-mediated recombination. Generalized Cre-mediated excisional recombination leads to surface expression of PLAP. Although recombination occurs in all tissues, PLAP expression is T-cell-specific owing to the CD2 promoter. Flow cytometric analysis of thymocytes from **a**, CD2-STOP-PLAP mice or **b**, CD2-STOP-PLAP x hCMV-Cre double transgenic mice shows T-cell-specific, recombination-mediated PLAP expression. Adjacent spleen sections from doubly transgenic mice were stained with **c**, Thy1.2 (blue) and B220 (brown) or **d**, PLAP. The entire splenic T-cell areas were stained by PLAP (**d**).

Fig. 1a) and the other has a T-cell-specific CD2 promoter¹¹ driving a Cre recombinase substrate that permanently expresses PLAP after recombination has occurred. In the CD2-STOP-PLAP transgene (Fig. 1b), expression of the downstream PLAP reporter gene is prevented by a *loxP*-flanked transcriptional and translational STOP fragment⁸. Following Cre-mediated recombination at the *loxP* sites, the STOP fragment is excised (Fig. 1b) and the PLAP reporter gene is positioned adjacent to, and under the control of, the CD2 promoter, which is transcriptionally active at all stages of T-cell differentiation^{12,13}.

We tested the cell-surface expression of PLAP following Cre-mediated recombination in multiple CD2-STOP-PLAP transgenic lines by mating each line with hCMV-Cre transgenic mice¹⁴ (human cytomegalovirus enhancer/promoter driving the *cre* gene). In hCMV-Cre x CD2-STOP-PLAP double transgenic mice, recombination occurred in all tissues, but PLAP was expressed solely in T cells owing to the specificity of the CD2 promoter. Fluorescence-activated cell sorting (FACS) analysis of thymocytes from CD2-STOP-PLAP and CD2-STOP-PLAP x hCMV-Cre

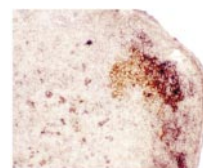


Figure 3 Expression of granzyme-B-promoter-driven Cre recombinase in activated T cells. *In vivo*, *cre* expression is observed in T cells of constitutively activated germinal centres in the Peyer's patch. Cre recombinase expression was visualized immunohistochemically using anti-Cre antibodies and germinal centres were identified by reactivity with the lectin peanut agglutinin (PNA). A representative PNA⁺ (brown) Peyer's patch germinal centre with *cre*⁺ (black) T cells is shown.

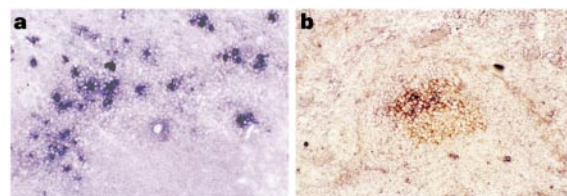


Figure 4 PLAP expression in CD4⁺ T cells responding to KLH *in vivo*. Doubly transgenic mice (CD2-STOP-PLAP x granzyme-B-Cre) were immunized with a single intraperitoneal immunization with 100 µg KLH in Ribi adjuvant. Frozen spleen sections of mice immunized 7 and 12 days earlier were stained with anti-PLAP antibodies. **a**, Clusters of PLAP⁺ T cells are seen in the T zones at day 7. **b**, PLAP⁺ T cells (black) in the PNA⁺ (brown) germinal centres at day 12.

transgenic mice demonstrated that before recombination in singly transgenic mice, PLAP was not expressed (Fig. 2a), whereas in the doubly transgenic mice virtually all of the thymic T cells expressed PLAP (Fig. 2b). In the spleens of these doubly transgenic mice, all of the T cells expressed PLAP (Fig. 2d). We generated three independent granzyme-B-driven *cre* transgenic lines. *In vivo*, *cre* was expressed specifically in activated CD4⁺ T cells in the Peyer's patch germinal centres (Fig. 3); activated CD4⁺ T cells are known to exist in germinal centres¹⁵. Spleen sections of granzyme-B-Cre transgenic mice immunized with keyhole limpet haemocyanin (KLH) also showed *cre* expression in the responding CD4⁺ T cells (data not shown).

We analysed doubly transgenic mice (CD2-STOP-PLAP x granzyme-B-Cre) for activation-induced expression of PLAP in CD4⁺ T cells by immunizing them with 100 µg of KLH. As in previous studies¹⁶⁻¹⁸, clusters of T cells expressing PLAP and CD4 were seen in the T zones at days 3, 5 and 7 after immunization (Fig. 4a). The size of these responding T-cell clusters peaked at day 5; by day 12, there were PLAP⁺ T cells in the germinal centres as well

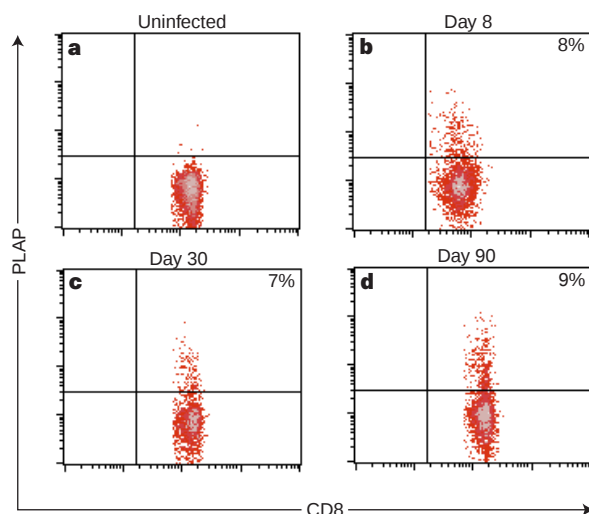


Figure 5 CD8⁺ T-cell response to LCMV infection in CD2-STOP-PLAP $\times$ granzyme-B-Cre mice. Mice were immunized with a single intraperitoneal injection of 5×10^4 PFU of the Armstrong strain of LCMV. Splenocytes of uninfected (**a**) and infected doubly transgenic mice were analysed by flow cytometry at 8 (**b**), 30 (**c**) and 90 (**d**) days post infection.

(Fig. 4b). We used double immunofluorescence confocal microscopy to show that PLAP⁺ cells expressed CD4, CD3, TcR $\alpha\beta$ and CD2 (data not shown). In contrast PLAP expression was not observed in the spleens of unimmunized double transgenic littermate controls.

To analyse acute and memory CD8⁺ T-cell responses we used the well-characterized lymphocytic choriomeningitis virus (LCMV) infection model^{19–24}. This natural mouse pathogen causes massive proliferation of CD8⁺ cells; mice clear the virus within 10 days, and LCMV-specific cytotoxic T cells persist for more than a year in increased numbers^{2,25}. CD2-STOP-PLAP $\times$ granzyme-B-Cre double transgenic mice were infected with LCMV by a single intraperitoneal injection of 5×10^4 plaque-forming units (PFU) of the Armstrong strain of LCMV, and their spleens were analysed at 8, 30, 60 and 90 days after infection. In control, uninfected doubly transgenic mice, all of the T cells were PLAP[–] by FACS (Fig. 5a) and by immunohistochemistry (data not shown). In the acute phase of the infection, LCMV induced massive splenomegaly and CD8⁺ T-cell proliferation—reversing the CD4:CD8 ratio (data not shown)—and 8–10% of the CD8⁺ T cells (approximately $4\text{--}6 \times 10^6$) expressed PLAP (Fig. 5b). Following virus clearance, the spleen sizes returned to normal and, in the memory phase of the response (days 30 and 90), 7–9% of the CD8⁺ T cells (approximately $1\text{--}2 \times 10^6$ cells) were PLAP⁺ (Fig. 5c, d), demonstrating that CD8⁺ T cells irreversibly marked in the acute LCMV response could be identified in the memory phase.

Although 70–90% of the CD8⁺ T cells showed the activated CD44^{hi}CD62L^{lo}Ly6C^{hi} phenotype (data not shown) eight days after infection, PLAP expression was observed in only 8–10% of the CD8⁺ T cells. During acute influenza A infection in mice, only a fraction of the activated T cells (~5%) scored positive in limiting dilution analysis (LDA)⁷. LDA also showed that only 1–5% of the responding CD8⁺ T cells were LCMV-specific at day 8 post-infection^{26,27}. This frequency of virus-specific CD8⁺ T cells, as judged by LDA, is in agreement with the frequency of T cells expressing PLAP and CD8 that we observe at day 8 after LCMV infection.

Experiments using direct staining with LCMV peptide major histocompatibility complex (MHC) tetramers²⁸ and/or quantifica-

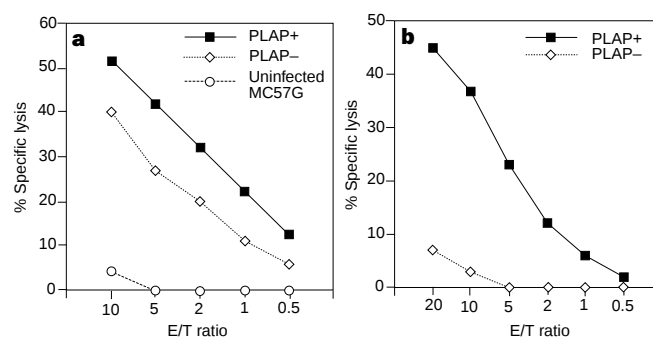


Figure 6 PLAP⁺ CD8⁺ T cells are LCMV-specific. **a**, Direct *ex vivo* killing of LCMV-infected target cells by PLAP⁺ and PLAP[–] CD8⁺ T cells from mice infected eight days earlier with LCMV. PLAP⁺ and PLAP[–] CD8⁺ T cells from CD2-STOP-PLAP $\times$ granzyme-B-Cre doubly transgenic mice were isolated by FACS and assayed for direct *ex vivo* cytotoxicity in a 6 h ⁵¹Cr-release assay. Shown is a representative direct *ex vivo* cytotoxicity assay of sorted PLAP⁺ and PLAP[–] CD8⁺ T cells (H-2^{b/a}) against uninfected or LCMV-infected MC57G (H-2^b) target cells. The percent specific lysis is plotted against effector:target (E/T) ratios used. **b**, Secondary bulk CTL assay. CD8⁺ PLAP⁺ and CD8⁺ PLAP[–] T cells were isolated by FACS from mice infected eight days earlier with LCMV. The sorted CD8⁺ PLAP⁺ and CD8⁺ PLAP[–] T cells (H-2^b) were stimulated *in vitro* with irradiated LCMV-infected peritoneal exudate cells (H-2^b), irradiated syngeneic splenocytes (H-2^b) and interleukin-2 (IL-2). After six days of incubation at 37 °C the cells were assayed for cytotoxicity against uninfected or LCMV-infected MC57G (H-2^b) target cells in a 6 h ⁵¹Cr-release assay.

tion of interferon- γ -producing T cells^{28,29} in an acute LCMV infection have shown that, contrary to the published LDA data, most of the responding CD8⁺ T cells (50–70%) at day 8 are LCMV specific. However, in our system, at day 8 after LCMV infection, only a subset (8–10%) of the CD8⁺ T cells is PLAP⁺. In the memory phase (>30 days), however, the frequencies of tetramer binding (10%)²⁸ and PLAP⁺ CD8⁺ T cells (Fig. 5c, d) are identical. Why then, in the acute phase of the LCMV response, does only a subset of the CD8⁺ T cells express PLAP? This difference was not due to the LCMV viral stocks used; we analysed responses to LCMV Arm viral stocks from three independent sources and found consistently that, at day 8, only 8–10% of the CD8⁺ T cells were PLAP⁺. Can both of these populations (PLAP⁺ and PLAP[–] CD8⁺ T cells) lyse virus-infected targets, as indicated by the high frequency (50–70%) of LCMV peptide/MHC tetramer-binding CD8⁺ T cells²⁸? To test this we used FACS to isolate CD8⁺PLAP⁺ and CD8⁺PLAP[–] T cells from mice infected eight days earlier with LCMV. The sorted cells were assayed for cytolytic activity in a direct *ex vivo* ⁵¹Cr-release cytotoxicity assay. PLAP⁺ and PLAP[–] CD8⁺ T-cell populations demonstrated comparable killing of LCMV-infected target cells (Fig. 6a). (The lytic units for the PLAP[–]CD8⁺ T cells are slightly lower because the sorted population also contains a small fraction (10–15%) of CD44^{lo} CD8⁺ T cells (data not shown) that do not participate in the response.) As both populations are LCMV-specific and cytolytic, what is the difference between the two? In the memory phase the frequency of PLAP⁺CD8⁺ T cells agrees with the frequency of tetramer-binding CD8⁺ T cells²⁸. Therefore, could the PLAP⁺ CD8⁺ T cells at day eight be precursors of memory T cells that have long-term growth potential? To test this, we performed secondary bulk cytotoxic T lymphocyte (CTL) assays on purified CD8⁺PLAP⁺ and CD8⁺PLAP[–] T cells isolated by FACS from double transgenic mice infected eight days earlier with LCMV by stimulating them with interleukin-2 (IL-2) and LCMV-infected peritoneal-exudate cells (PEC) from syngeneic mice. Following *in vitro* stimulation for 6 days at 37 °C, the ability of the stimulated cells to lyse LCMV-infected targets was assayed in a standard 6–8 h ⁵¹Cr-release assay. In the secondary bulk CTL assay (Fig. 6b), only CD8⁺ PLAP⁺ T cells could kill LCMV-infected targets. There appear to be two populations of effector CTLs in the acute LCMV response;

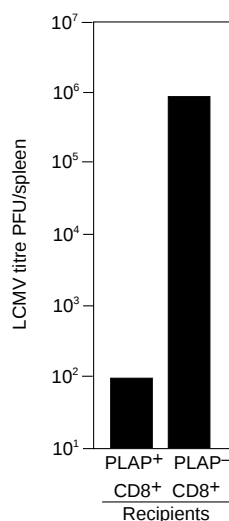


Figure 7 Adoptive transfer of CD8⁺ T cell memory. CD8⁺ PLAP⁺ and CD8⁺ PLAP⁻ were sorted by FACS from LCMV-immune (infected >1 month earlier) doubly transgenic mice. 2×10^5 sorted cells were adoptively transferred into naive recipients and challenged the next day with 2×10^6 PFU clone-13 LCMV, intravenously. Eight days later their spleens were removed and LCMV titres were determined as described³⁰. The mean LCMV titres (log₁₀) per spleen of mice that received CD8⁺ PLAP⁺ and CD8⁺ PLAP⁻ T cells are shown.

PLAP⁺ and PLAP⁻ CTLs both of which are capable of direct *ex vivo* cytolytic activity (Fig. 6a). However, only the PLAP⁺ CD8⁺ T cells are capable of cytolytic activity after long term *in vitro* stimulation (Fig. 6b). Thus, although most (50–70%) if the CD8⁺ T cells during an acute LCMV infection are LCMV specific, as shown by tetramer staining, only a small fraction (8–10% CD8⁺ PLAP⁺) is capable of long-term growth and survival.

To determine whether PLAP⁺ CD8⁺ T cells observed in the memory phase of the response (Fig. 5c, d) are LCMV-specific memory T cells, we conducted adoptive transfer experiments as described²⁵. Briefly, PLAP⁺ and PLAP⁻ CD8⁺ T cells from mice infected 1–2 months earlier were purified by cell sorting and adoptively transferred into naive syngeneic recipient mice, which were challenged with LCMV the next day. Eight days after the challenge, virus titres in the recipient spleens were determined by plaque assay on Vero cells³⁰. Mice that received 2×10^5 PLAP⁺ CD8⁺ T cells cleared the virus much more efficiently than mice that received 2×10^5 PLAP⁻ CD8⁺ T cells (LCMV titres were 4 logs lower) (Fig. 7). Thus, the PLAP⁺ CD8⁺ T cells seen in the memory phase are LCMV-specific memory T cells.

It has been difficult to identify memory T cells unambiguously because of the inability to distinguish between recently activated effector cells or memory precursor cells and true memory T cells^{4,26}. If the granzyme-B-promoter activity is extinguished in memory T cells, then it should be possible to distinguish between naive, activated and memory CD8⁺ T cells by looking at granzyme-B-promoter-driven Cre expression. After infection, memory CD8⁺ T cells should be PLAP⁺ cre⁻, whereas recently activated CD8⁺ T cells will be PLAP⁺ cre⁺ and naive T cells will be PLAP⁻ cre⁻. We analysed PLAP and cre expression in the spleens of doubly transgenic mice during the peak of the acute infection (day 8) and during the memory phase (day 60). During acute infection, PLAP⁺ CD8⁺ T cells could easily be identified in spleen sections (Fig. 8a). In the adjacent section, the same region contained roughly equal frequencies of cre⁺ CD8⁺ T cells (Fig. 8b); the ratio of the frequencies of PLAP⁺ and cre⁺ T cells in the same splenic areas in adjacent sections was 1:0.9. In contrast, in the memory stage, two months after LCMV infection, very few PLAP⁺ CD8⁺ T cells outside the germinal centres

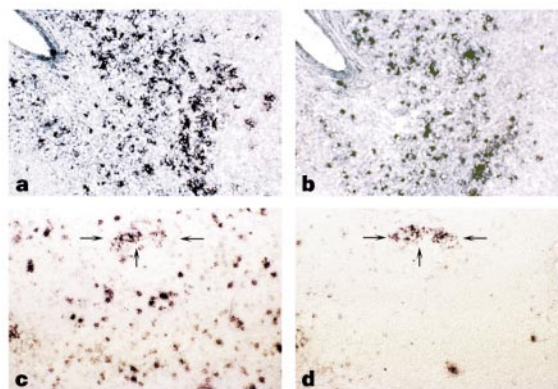


Figure 8 Memory T cells can be distinguished from recently activated T cells. Frozen spleen sections taken from CD2-STOP-PLAP × granzyme-B-Cre doubly transgenic mice during the acute LCMV-response phase (day 8; **a**, **b**) and memory phase (day 60; **c**, **d**) were analysed for PLAP (**a**, **c**) and cre expression (**b**, **d**). Spleen sections were stained with rabbit anti-PLAP antibodies (**a**, **c**) and the adjacent sections were stained with rabbit anti-Cre antibodies (**b**, **d**). The bound antibodies were visualized using a biotinylated anti-rabbit IgG secondary antibody followed by ABC-AP and NBT/BCIP for colour development. In the spleen sections stained for Cre, transgenic PLAP activity was eliminated using acid treatment (15% acetic acid, 5 min). At day 8, virtually all of the CD8⁺ T cells were both PLAP⁺ (**a**) and Cre⁺ (**b**), whereas in the memory phase most of the CD8⁺ T cells outside the germinal centres were PLAP⁺ (**c**) but Cre⁻ (**d**). The CD4⁺ T cells in the germinal centres (arrows) at day 60 were Cre⁺.

(Fig. 8c) were cre⁺ (Fig. 8d). As anticipated, the CD4⁺ T cells in the active germinal centres at day 60 were PLAP⁺ cre⁺ (Fig. 8d). This is consistent with published data demonstrating that activated CD4⁺ T cells persist in germinal centres¹⁵.

Our data show that memory T cells can be indelibly marked. The cells were proven to be virus-specific memory cells by adoptive transfer. In addition, the cells that will give rise to the memory CD8⁺ pool can be identified in the primary response by expression of PLAP promoted by truncated granzyme-B promoter. This was not predicted and we do not know why the marking distinguishes active short-lived effector CTLs (PLAP⁻ CD8⁺) from the PLAP⁺ CD8⁺ CTLs that will also give rise to the memory pool. Nonetheless, we show that these are distinct pools and the methodology allows us to separate the two kinds of cell. In uninfected mice, 5–10% of T cells are CD44^{hi}, presumably because of environmental antigens, but none of these cells are PLAP⁺. This suggests that the activation requirements for the CD44^{hi} phenotype and the truncated-granzyme-B-promoted induction into the memory pool are distinct. We have also shown that memory T cells can be distinguished from activated, cytolytic, pre-memory T cells by differential Cre expression.

Our model system allows indelible marking of activated CD4⁺ and CD8⁺ T cells. As *a priori* knowledge of immunodominant peptides is not needed, our system can be used to study T-cell responses against any pathogen. The model system can be used for visualizing and fate-mapping activated T cells involved in transplantation as well as autoimmune diseases. The ability to identify and purify responding activated T cells and memory T cells makes molecular, genetic and biochemical analyses possible and allows the analysis of the generation, maintenance and regulation of T-cell memory. The capability to directly quantify the memory T-cell pool could also make this mouse model system valuable for screening candidate vaccines.

Methods

DNA constructs and transgenic mice. The PLAP coding sequence was cloned into pBS302 and the NotI lox-STOP-loxP-PLAP fragment was blunted and cloned into the SmaI site of VahCD2 (ref. 11) to generate the CD2-STOP-PLAP construct. The (NotI/XbaI) transgenic fragment was used for transgenesis and

six independent founder mice (FvB/N × C57BL/6) were obtained from 18 pups that were born. The *cre* gene was modified to contain the nuclear localization sequence of the SV40 large T antigen by PCR and cloned into the *Bam*HI site of p687 (ref. 10) and the *Eco*RI fragment was used to generate transgenic mice. Of the ten pups that were born, four carried the transgene. The founder mice were bred to generate transgenic lines.

Immunization and infection. Cohorts of transgenic and non-transgenic littermate control mice were immunized intraperitoneally with 100 µg KLH emulsified in Ribi adjuvant. For LCMV infection, mice were injected intraperitoneally with 5×10^4 PFU of the Armstrong strain of LCMV. Spleens were collected and analysed by frozen sectioning and FACS analysis at 8, 30, 60 and 90 days post-infection.

Cytotoxicity assay. CTL assays were performed as described². Sorted CD8⁺ PLAP⁺ and PLAP⁻ T cells (H-2^{b/q}, H-2^b or H-2^d) were plated in triplicate to achieve the desired effector/target ratio along with ⁵¹Cr-labelled uninfected or LCMV-infected target cells. MC57G (H-2^b) and CRL-2110 (H-2^d) (ATCC) were used as target cells. Following 6 h incubation at 37°C the supernatant was collected and counted. Percent specific lysis was calculated as $100 \times [(experimental\ c.p.m. - spontaneous\ c.p.m.) / (maximum\ release - spontaneous\ release)]$.

Secondary bulk CTL assay. CD8⁺ PLAP⁺ and PLAP⁻ T cells were purified by FACS from mice infected eight days earlier with LCMV. 10^6 sorted cells (CD8⁺ PLAP⁺ and PLAP⁻) were co-cultured with 8×10^5 irradiated (1,200 rad) syngeneic spleen cells, 2×10^5 irradiated LCMV-infected, thioglycollate-induced PEC and IL-2 (20 units) in a 200 µl volume. After six days of incubation at 37°C, their cytotoxicity was tested in a 6–8-h ⁵¹Cr-release assay.

Adoptive transfer. Adoptive transfer and viral challenge were done as described²⁵. Briefly, CD8⁺ PLAP⁺ and PLAP⁻ T cells were isolated by FACS from LCMV-immune doubly transgenic mice (infected >1 month earlier with LCMVArm). The sorted cells were injected into the lateral tail vein of syngeneic naive recipients that had been irradiated 1 day earlier (600 rad). Each recipient received 2×10^5 PLAP⁺ or PLAP⁻ CD8⁺ T cells. The day after adoptive transfer, recipient mice were challenged with a single i.v. injection of 2×10^6 PFU of clone 13 LCMV. Eight days after the viral challenge, the spleens were removed, homogenized and titred for LCMV by plaquing on Vero cells³⁰.

Received 25 January; accepted 9 April 1999.

- Hou, S., Hyland, L., Ryan, K. W., Portner, A. & Doherty, P. C. Virus-specific CD8⁺ T-cell memory determined by cloning burst size. *Nature* **369**, 652–654 (1994).
- Welsh, R. M. *et al.* Alpha beta and gamma delta T-cell networks and their roles in natural resistance to viral infections. *Immunol. Rev.* **159**, 79–93 (1997).
- Kundig, T. M. *et al.* On T cell memory: arguments for antigen dependence. *Immunol. Rev.* **150**, 63–90 (1996).
- Mullbacher, A. & Flynn, K. Aspects of cytotoxic T cell memory. *Immunol. Rev.* **150**, 113–127 (1996).
- Tough, D. F. & Sprent, J. Lifespan of lymphocytes. *Immunol. Rev.* **14**, 1–12 (1995).
- Swain, S. L. *et al.* From naive to memory T cells. *Immunol. Rev.* **150**, 143–167 (1996).
- Doherty, P. C., Topham, D. J. & Tripp, R. A. Establishment and persistence of virus-specific CD4⁺ and CD8⁺ T cell memory. *Immunol. Rev.* **150**, 23–44 (1996).
- Lasko, M. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6232–6236 (1992).
- Cepko, C., Ryder, E. F., Austin, C. P., Walsh, C. & Fekete, D. M. Lineage analysis using retrovirus vectors. *Methods Enzymol.* **254**, 387–419 (1995).
- Hanson, R. D., Sclar, G. M., Kanagawa, O. & Ley, T. J. The 5'-flanking region of the human CGL-1/granzyme B gene targets expression of a reporter gene to activated T-lymphocytes in transgenic mice. *J. Biol. Chem.* **266**, 24433–24438 (1991).
- Zhumabekov, T., Corbella, P., Tolaini, M. & Kioussis, D. Improved version of a human CD2 minigene based vector for T cell-specific expression in transgenic mice. *J. Biol. Chem.* **266**, 24433–24438 (1991).
- Williams, A. F., Barclay, A. N., Clark, S. J., Paterson, D. J. & Willis, A. C. Similarities in sequences and cellular expression between rat CD2 and CD4 antigens. *J. Exp. Med.* **165**, 368–380 (1987).
- Wotton, D., Flanagan, B. F. & Owen, M. J. Chromatin configuration of the human CD2 gene locus during T-cell development. *Proc. Natl Acad. Sci. USA* **86**, 4195–4199 (1989).
- Nagy, A. *et al.* Multipurpose gene alterations from a single targeting vector: dissecting the role of N-myc in development. *Curr. Biol.* **8**, 661–664 (1998).
- Castan, J., Tenner, R. K., Racz, P., Fleischer, B. & Broker, B. M. Accumulation of CTLA-4 expressing T lymphocytes in the germinal centres of human lymphoid tissues. *Immunology* **90**, 265–271 (1997).
- Zheng, B., Han, S., Zhu, Q., Goldsby, R. & Kelsae, G. Alternative pathways for the selection of antigen-specific peripheral T cells. *Nature* **384**, 263–266 (1996).
- Fuller, K. A., Kanagawa, O. & Naim, M. H. T cells within germinal centres are specific for the immunizing antigen. *J. Immunol.* **151**, 4505–4512 (1993).
- Gulbranson, J. A. & MacLennan, I. Sequential antigen-specific growth of T cells in the T zones and follicles in response to pigeon cytochrome c. *Eur. J. Immunol.* **26**, 1830–1837 (1996).
- Ahmed, R., Byrne, J. A. & Oldstone, M. B. Virus specificity of cytotoxic T lymphocytes generated during acute lymphocytic choriomeningitis virus infection: role of the H-2 region in determining cross-reactivity for different lymphocytic choriomeningitis virus strains. *J. Virol.* **51**, 34–41 (1984).
- Jamieson, B. D. & Ahmed, R. T cell memory. Long-term persistence of virus-specific cytotoxic T cells. *J. Exp. Med.* **169**, 1993–2005 (1989).
- Selin, L. K., Nahill, S. R. & Welsh, R. M. Cross-reactivities in memory cytotoxic T lymphocyte recognition of heterologous viruses. *J. Exp. Med.* **179**, 1933–1943 (1994).
- Zimmerman, C., Brduscha, R. K., Blaser, C., Zinkernagel, R. M. & Pircher, H. Visualization,

- characterization, and turnover of CD8⁺ memory T cells in virus-infected hosts. *J. Exp. Med.* **183**, 1367–1375 (1996).
- Nahill, S. R. & Welsh, R. M. High frequency of cross-reactive cytotoxic T lymphocytes elicited during the virus-induced polyclonal cytotoxic T lymphocyte response. *J. Exp. Med.* **177**, 317–327 (1993).
- Yang, H. Y., Dundon, P. L., Nahill, S. R. & Welsh, R. M. Virus-induced polyclonal cytotoxic T lymphocyte stimulation. *J. Immunol.* **142**, 1710–1718 (1989).
- Lau, L. L., Jamieson, B. D., Somasundaram, T. & Ahmed, R. Cytotoxic T-cell memory without antigen. *Nature* **369**, 648–652 (1994).
- Ahmed, R. & Gray, D. Immunological memory and protective immunity: understanding their relation. *Science* **272**, 54–60 (1996).
- Zinkernagel, R. M. *et al.* On immunological memory. *Annu. Rev. Immunol.* **14**, 333–367 (1996).
- Murali, K. K. *et al.* Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
- Butz, E. A. & Bevan, M. J. Massive expansion of antigen-specific CD8⁺ T cells during an acute virus infection. *Immunity* **8**, 167–175 (1998).
- Ahmed, R., Salmi, A., Butler, L. D., Chiller, J. M. & Oldstone, M. B. A. Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice: role in suppression of cytotoxic T lymphocyte response and viral persistence. *J. Exp. Med.* **160**, 521–540 (1984).

Acknowledgements. We thank C. Roman, I. Stancovski, S. Cherry, K. Alexandropoulos, M. Scott, G. Kelsae, R. Welsh, L. Selin, R. Ahmed and members of the Baltimore laboratory for useful discussions. We also thank T. Ley for p687 plasmid; R. Welsh, R. Ahmed and J. Carlos de la Torre for LCM viral stocks; D. Kioussis for the VahCD2 transgenic construct; M. Brach for pBS185 and pBS302 plasmids; and A. Nagy for the hCMV-Cre transgenic mice. J.J. Thanks H. Rayburn and W. Sha for teaching him pronuclear injection and transgenesis, and M. Scott for help with adoptive transfers. J.J. was a recipient of a postdoctoral fellowship from the Helen Hay Whitney foundation and is currently a special fellow of the Leukemia Society of America.

Correspondence and requests for materials should be addressed to D.B. (e-mail: baltimo@caltech.edu).

Regulation of endothelium-derived nitric oxide production by the protein kinase Akt

David Fulton*, Jean-Philippe Gratton*, Timothy J. McCabe*, Jason Fontana*, Yasushi Fujio†, Kenneth Walsh‡, Thomas F. Franke‡, Andreas Papapetropoulos* & William C. Sessa*

* Department of Pharmacology and Molecular Cardiology Program, Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, Connecticut 06536, USA

† Cardiovascular Research, St. Elizabeth's Medical Center, Boston, Massachusetts 02135, USA

‡ Department of Pharmacology, Columbia University, New York, New York 10032, USA

Endothelial nitric oxide synthase (eNOS) is the nitric oxide synthase isoform responsible for maintaining systemic blood pressure, vascular remodelling and angiogenesis^{1–4}. eNOS is phosphorylated in response to various forms of cellular stimulation^{5–7}, but the role of phosphorylation in the regulation of nitric oxide (NO) production and the kinase(s) responsible are not known. Here we show that the serine/threonine protein kinase Akt (protein kinase B) can directly phosphorylate eNOS on serine 1179 and activate the enzyme, leading to NO production, whereas mutant eNOS (S1179A) is resistant to phosphorylation and activation by Akt. Moreover, using adenovirus-mediated gene transfer, activated Akt increases basal NO release from endothelial cells, and activation-deficient Akt attenuates NO production stimulated by vascular endothelial growth factor. Thus, eNOS is a newly described Akt substrate linking signal transduction by Akt to the release of the gaseous second messenger NO.

The Akt proto-oncogene is an important regulator of various cellular processes including glucose metabolism and cell survival^{8,9}. Activation of receptor tyrosine kinases and G-protein-coupled receptors, and stimulation of cells by mechanical forces, can lead to the phosphorylation and activation of Akt^{10–12}. Akt can then phosphorylate substrates such as glycogen synthase kinase-3, Bad and caspase-9, resulting in protein inactivation, or 6-phospho-fructo-2-kinase, resulting in protein activation^{10,13}. The relationships between activation of Akt, its downstream effectors and the production of soluble second messengers are not well understood.

eNOS is the nitric oxide synthase (NOS) isoform that produces endothelium-derived NO. Treatment of endothelial cells with vascular endothelial growth factor (VEGF) or insulin stimulates the production of NO by a phosphatidylinositol-3-OH-kinase (PI(3)K)-dependent mechanism^{14,15}. Wortmannin and LY294002, two structurally dissimilar inhibitors of PI(3)K, partially block NO release. VEGF also stimulates the Ras pathway; inhibition of Ras

signalling blocks extracellular-signal-related-kinase (ERK)-1/2 activation, but not VEGF-stimulated NO production, indicating that there may be a link between growth factor signalling through PI(3)K and eNOS¹⁶.

To investigate whether Akt, a downstream effector of PI(3)K, could directly influence the production of NO, COS-7 cells (which do not express NOS) were co-transfected with eNOS and wild-type

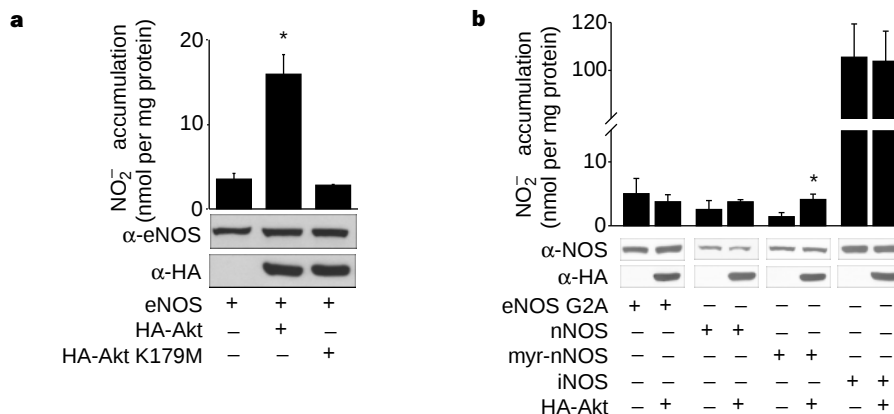


Figure 1 Wild-type Akt, but not kinase-inactive Akt, increases NO release from cells expressing membrane-associated eNOS. **a**, COS cells were transfected with plasmids for eNOS, with or without Akt or kinase-inactive Akt (K179M). The production of NO (assayed as NO₂⁻) was determined by chemiluminescence. **b**, COS cells were transfected with various NOS plasmids as above. In both **a** and

b, values for NO₂⁻ production were subtracted from levels obtained from cells transfected with the β -galactosidase cDNA only. The insets show the expression of proteins in total cell lysates. Data were mean $\pm$ s.e.m., $n = 3-7$ experiments; asterisk denotes $P < 0.05$.

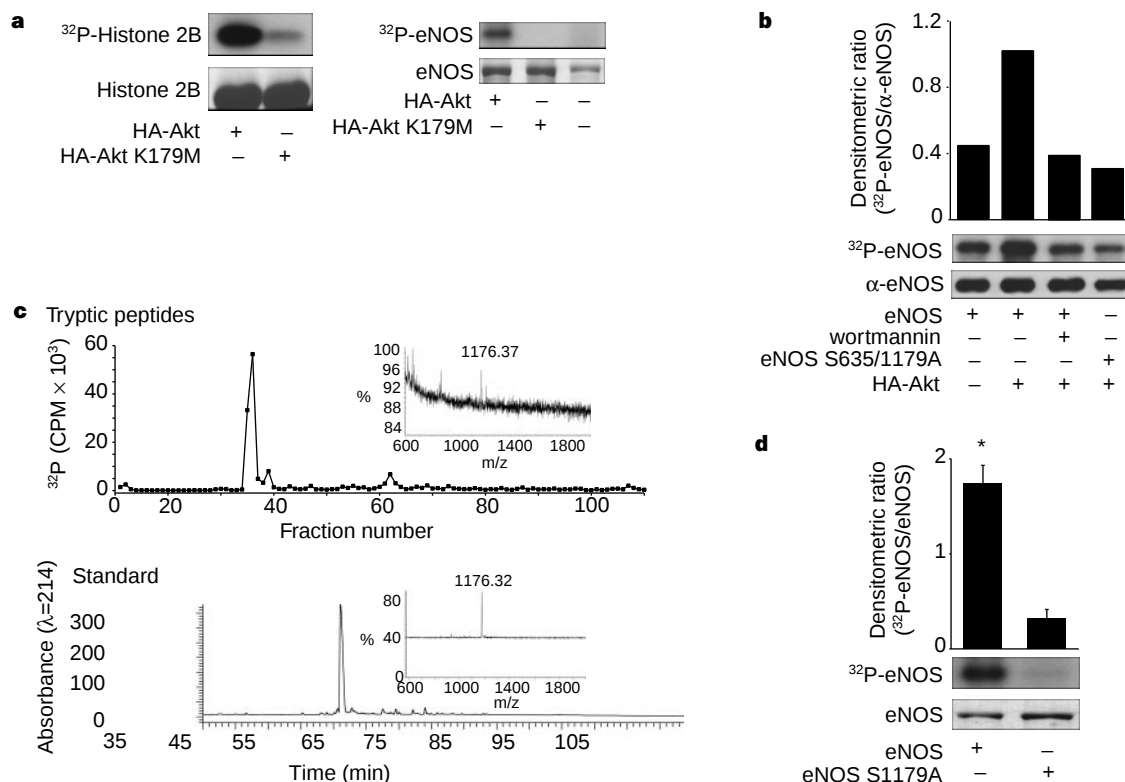


Figure 2 Phosphorylation of eNOS by active Akt *in vitro* and *in vivo*. **a**, COS cells were transfected with HA-Akt or HA-Akt(K179M) and lysates were immunoprecipitated and placed into an *in vitro* kinase reaction with histone 2B (25 μ g) or recombinant eNOS (3 μ g) as substrates; Top panel: incorporation of ³²P into the substrates; Bottom panel: amount of substrate by Coomassie staining of the gel. **b**, ³²P-labelled wild-type or double-serine mutant of eNOS (eNOS S635/1179A) was affinity purified from transfected COS cells and subjected to autoradiography (upper panel) or western blotting (lower panel). The histogram shows the relative amount of labelled protein to the amount of immunoreactive eNOS in the gel. **c**, Labelled eNOS was digested with trypsin and peptides separated by RP-

HPLC. The upper panel shows a predominant labelled tryptic peptide that co-migrates with the unlabelled synthetic phosphopeptide standard (bottom panel). Insets: linear mode MS of labelled peptide (top) and phosphopeptide standard showing identical mass ions (bottom). **d**, Recombinant wild-type eNOS or eNOS S1179A were purified and equal amounts (2.4 μ g) placed into an *in vitro* kinase reaction with recombinant Akt (see Methods). Top panel: the incorporation of ³²P into eNOS; bottom panel; amount of substrate by Coomassie staining of the gel. Histogram ($n = 3$) reflects the relative amount of labelled eNOS to the mass of eNOS (Coomassie) in the *in vitro* kinase reaction.

Akt (HA-Akt) or kinase-inactive Akt (HA-Akt K179M), and the accumulation of nitrite (NO_2^-) was measured by NO-specific chemiluminescence. Transfection of eNOS results in increased NO_2^- accumulation, which is markedly enhanced by co-transfection of wild-type Akt, but not the kinase-inactive variant (Fig. 1a). Identical results were obtained using cyclic GMP (cGMP) as a bioassay for biologically active NO. Transfection of a constitutively active form of Akt (myr-Akt) increases cGMP accumulation (assayed in COS cells) from 5.5 ± 0.8 to 11.6 ± 0.9 pmol cGMP per mg protein (in cells transfected with eNOS alone or eNOS with myr-Akt, respectively) whereas the kinase-inactive Akt did not influence cGMP accumulation (5.8 ± 0.8 pmol cGMP, $n = 4$ experiments). Under these experimental conditions, Akt was catalytically active, as determined by western blotting with a phospho-Akt-specific antibody (which recognizes serine 473; data not shown) and Akt activity assays (see Fig. 2a). Equal levels of eNOS and Akt were expressed in COS cell lysates (Fig. 1), indicating that Akt modulates eNOS, thereby increasing NO production under basal conditions.

eNOS is a dually acylated peripheral membrane protein that targets into the Golgi region and plasma membrane of endothelial cells^{17–19}; compartmentalization is required for efficient production of NO in response to agonist challenge^{20–22}. To test whether eNOS activation by Akt requires membrane compartmentalization, we co-transfected COS-7 cells with complementary DNAs for Akt and a myristoylation, palmitoylation-defective mutant of eNOS (G2A eNOS); we then quantified the release of NO. Akt did not activate the non-acylated form of eNOS (Fig. 1b), indicating that compartmentalization of both proteins to the membrane is required for their functional interaction¹⁰. Next, we tested whether Akt could activate two structurally similar but distinct soluble NOS isoforms, neuronal NOS and inducible NOS (nNOS and iNOS, respectively). Co-transfection of Akt with nNOS and iNOS did not result in a further increase in NO release, demonstrating the specificity of Akt for eNOS. However, the addition of an N-myristoylation site to nNOS, to enhance its interactions with biological membranes, results in Akt stimulation of nNOS in a manner analogous to that seen with eNOS, indicating that both isoforms may be susceptible to activation by Akt when anchored to the membrane.

The above data indicate that Akt, perhaps by phosphorylation of eNOS, can modulate NO release from intact cells. Indeed, two putative Akt phosphorylation motifs (RXRXXS/T) are present in

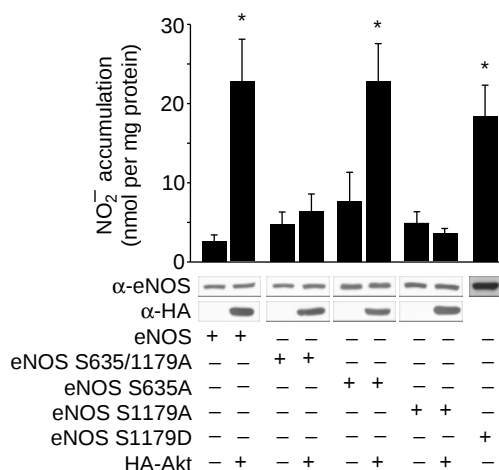


Figure 3 Evidence that Ser 1179 is functionally important for Akt-stimulated NO release. COS cells were transfected with plasmids for wild-type eNOS or eNOS mutants, in the absence or presence of Akt, and the expression of the proteins and production of NO (assayed as NO_2^-) were determined. Constructs with the S1179A mutation were not activated by Akt; the S1179D mutation resulted in a gain of function. Data are mean $\pm$ s.e.m. of 4–7 experiments; asterisks represent significant differences ($P < 0.05$).

eNOS (serines 635 and 1179 in bovine eNOS; serines 633 and 1177 in human eNOS) and one motif is present in nNOS (serines 1412 in rat and 1415 in human nNOS), with no obvious motifs found in iNOS. To test whether eNOS is a potential substrate for Akt phosphorylation *in vitro*, we transfected COS cells with HA-Akt or HA-Akt (K179M), and assessed kinase activity using recombinant eNOS as a substrate. The active kinase phosphorylates histone 2B and eNOS (Fig. 2a; 69.3 ± 2.9 and 115.4 ± 3.8 pmol of ATP per nmol substrate, respectively; $n = 3$), whereas the inactive Akt did not significantly increase histone or eNOS phosphorylation. To investigate whether the putative Akt phosphorylation sites in eNOS were responsible for the incorporation of ^{32}P , the two serines were mutated to alanine residues and the ability of Akt to stimulate wild-type and mutant eNOS phosphorylation was examined in intact COS cells. We labelled transfected cells with ^{32}P -orthophosphate and partially purified eNOS by ADP-sepharose affinity chromatography, and quantified the phosphorylation state and protein levels. Co-expression of Akt results in a twofold enhancement in the phosphorylation of eNOS relative to non-stimulated cells (Fig. 2b). Pretreatment of eNOS/Akt-transfected cells with wortmannin abolished the Akt-induced increase in phosphorylation. Moreover, mutation of serines 635 and 1179 to alanine residues abolished Akt-dependent phosphorylation of eNOS, indicating that these residues could serve as potential phosphorylation sites in intact cells. To directly identify the residues phosphorylated by Akt, we incubated wild-type eNOS with immunopurified Akt and determined the sites of phosphorylation by high-performance liquid chromatography followed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). The primary ^{32}P -labelled tryptic phos-

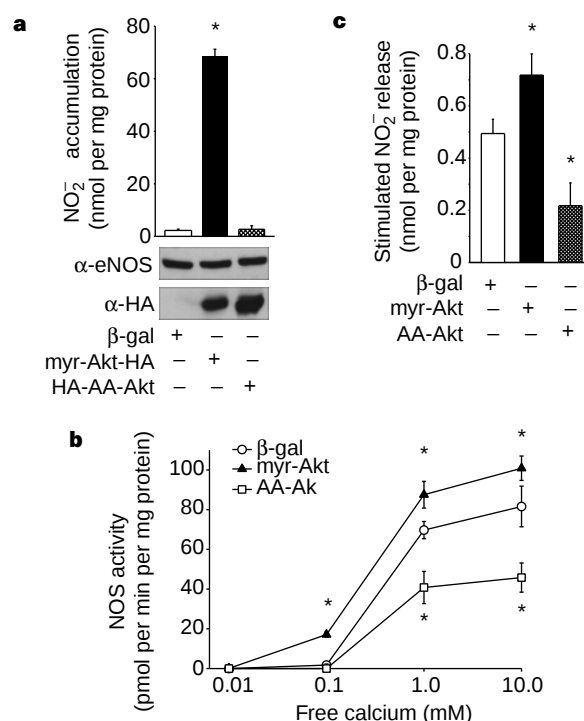


Figure 4 Akt regulates the basal and stimulated production of NO in endothelial cells. **a**, BLMVEC were infected with adenoviral constructs (β -gal as control, myr-Akt and AA-Akt) and the amount of NO_2^- produced over 24 h was determined ($n = 3$). Inset: expression of eNOS and Akt. **b**, Lysates from adenovirus-infected BLMVEC were examined for NOS activity. Equal amounts of protein (50 μg) were incubated with various concentrations of free calcium, and the NOS activity was determined ($n = 3$ experiments). **c**, BLMVEC were infected with adenoviruses as above followed by stimulation with VEGF (40 ng per ml) for 30 min and NO_2^- release was quantified by chemiluminescence. Data are presented as VEGF-stimulated release of NO_2^- after subtraction of basal levels; mean $\pm$ s.e.m., $n = 4$; asterisks represent significant differences ($P < 0.05$).

phopeptide co-elutes with a synthetic phosphopeptide (amino acids 1177–1185 with phosphoserine at position 1179) and has the identical mass ion as determined by linear mode MS (Fig. 2c). Using reflectron mode MALDI-MS monitoring, both the labelled tryptic peptide and the standard phosphopeptide demonstrated a loss of H_3PO_4 , indicating that the tryptic peptide was phosphorylated. In addition, mutation of Ser 1179 to alanine markedly reduces Akt-dependent phosphorylation of eNOS compared with the wild-type protein (Fig. 2d). Identical results were obtained using peptides (amino acids 1174–1194) derived from wild-type or eNOS S1179A as substrates for recombinant Akt (the wild-type peptide incorporated 24.6 ± 3.7 nmol phosphate per mg compared to the alanine mutant peptide, which incorporated 0.22 ± 0.02 nmol phosphate per mg; $n = 5$). These data show that eNOS is a substrate for Akt and that the primary site of phosphorylation is serine 1179.

Next we examined the functional significance of the putative Akt phosphorylation site at Ser 635 and the identified site at Ser 1179. Transfection of COS cells with the double mutant eNOS S635/1179A abolishes Akt-dependent NO release. Mutation of Ser 635 to alanine did not attenuate NO release, whereas eNOS S1179A abolishes Akt-dependent activation of eNOS (Fig. 3). These results indicate that Ser 1179 is functionally important for NO release. Mutation of Ser 1179 into aspartic acid (eNOS S1179D), to substitute for the negative charge afforded by the addition of phosphate, partially mimics the activation state induced by Akt. All site-directed mutants were amply expressed (see inset western blots) and retained NOS catalytic activity in cell lysates (in COS cells transfected with eNOS only, NOS activity was 85.3 ± 27.0 , 71.9 ± 2.9 , 80.8 ± 23.2 and 131.8 ± 36.7 pmol L-citrulline generated per mg protein from lysates of COS cells expressing wild-type, S1179A, S635/1179A and S1179D eNOS, respectively; $n = 3$ experiments).

To test whether Akt mediates NO release from endothelial cells, we infected bovine lung microvascular endothelial cells (BLMVEC) with adenoviruses expressing activated Akt (myr-Akt), activation-deficient Akt (AA-Akt) or β -galactosidase (as a control), and measured the accumulation of NO. Myr-Akt stimulates the basal production of NO from BLMVEC, whereas cells infected with β -galactosidase or activation-deficient Akt released small amounts of NO that were close to the limits of detection (Fig. 4a). These results, in conjunction with the results in COS cells, indicate that phosphorylation of eNOS by Akt is sufficient to regulate NO production at resting levels of calcium. Indeed, NOS activity measured in lysates from myr-Akt-infected BLMVEC demonstrates that the sensitivity of the enzyme to activation by calcium, assayed at a fixed calmodulin concentration, is enhanced relative to NOS activity seen in BLMVEC infected with the β -galactosidase virus (Fig. 4b). The calcium sensitivity of NOS activity in cells infected with activation-deficient Akt was greatly suppressed relative to both myr-Akt and β -galactosidase-infected cells.

As mentioned above, treatment of endothelial cells with VEGF activates Akt²³ and the release of NO through a mechanism partially blocked by inhibitors of PI(3)K¹⁵. Indeed, treatment of quiescent endothelial cells with VEGF induces Akt and eNOS phosphorylation (approximately twofold) concomitant with NO release (see Supplementary Information). To examine the functional link between VEGF as an agonist for NO release and Akt activation, BLMVEC were infected with adenoviruses for myr-Akt, AA-Akt or β -galactosidase, and VEGF-stimulated NO release was quantified. Infection of endothelial cells with myr-Akt enhances VEGF-driven NO production, whereas AA-Akt attenuates NO release (Fig. 4c). These results indicate that Akt may participate in the signal transduction events required for both basal and stimulated NO production in endothelial cells.

Our data show that Akt can phosphorylate eNOS on Ser 1179 and that phosphorylation enhances the ability of the enzyme to generate NO. Our results indicate that stimuli that promote NO release through a PI(3)K-dependent mechanism, independent of eleva-

tions in intracellular calcium (insulin and shear stress^{14,24,25}), will stimulate Akt, thus permitting the phosphorylation and activation of eNOS at resting levels of calcium (Figs 1, 4). This can occur contemporaneously with the recruitment of the NOS regulatory protein Hsp90²⁶, perhaps by stabilizing calmodulin previously bound to eNOS, thus resulting in NO release. In the case of G-protein-coupled receptors or receptor tyrosine kinases that mobilize intracellular calcium, we propose that agonist-dependent stimulation of PI(3)K¹¹ and/or activation of calmodulin-dependent protein kinase kinase (CaMKK)²⁷ will activate Akt to phosphorylate eNOS concomitantly with the increase in cytoplasmic calcium. The calcium-dependent activation of calmodulin will stimulate CaMKK and the recruitment of calmodulin and perhaps Hsp90 to eNOS, facilitating rapid eNOS activation and the burst-like release of NO. How phosphorylation of eNOS by Akt enhances NO release is not known, but it is likely to be related to changes in the sensitivity of the enzyme to calcium-activated calmodulin. This may occur by the introduction of a negative charge, either by phosphate or aspartate, thereby 'opening up' the structure (perhaps by removing the auto-inhibitory loop of eNOS²⁸) and thus permitting activated calmodulin binding at lower calcium concentrations. Thus, regulation of NO production by Akt-dependent phosphorylation of eNOS may provide a novel therapeutic target for the design of drugs aimed at improving endothelial function in cardiovascular diseases associated with dysfunction in the synthesis or biological activity of NO, such as hypertension, atherosclerosis, heart failure and diabetes.

Note added in proof: While in review, a recent paper has demonstrated that AMP-activated kinase can phosphorylate human eNOS on Ser 1177, *in vitro* (Chen *et al.* *FEBS Lett.*, **443**, 285–289 (1999)). □

Methods

Cell transfections. The bovine eNOS, human iNOS and rat nNOS cDNAs in pcDNA3 and HA-tagged wild-type Akt, Akt (K179M) and myr-Akt in pCMV6 were generated by standard cloning methods. The myr-nNOS in pcDNA3 was generated by PCR, incorporating a new amino terminus containing the eNOS N-myristoylation consensus site (MGNLKSVG) fused in frame to the second amino acid of the nNOS coding sequence. In preliminary experiments in COS cells, this construct was N-myristoylated based on incorporation of ³H-myristic acid (unlike native nNOS) and resulted in approximately 60% of the total protein being targeted into the membrane fraction of cells (only 5–10% of nNOS was membrane associated in COS cells). The putative Akt phosphorylation sites in eNOS were mutated using the Quick Change site-directed mutagenesis kit (Stratagene) according to the manufacturers instructions. All mutants were verified by DNA sequencing. We plated COS-7 cells (100 mm dish) and transfected them with the NOS (7.5–30 μ g) and Akt (1 μ g) plasmids using calcium phosphate. To balance all transfections, the expression vector for β -galactosidase cDNA was used. 24–48 h after transfection, the expression of appropriate proteins (40–80 μ g) were confirmed by western blot analysis using eNOS mAb (9D10, Zymed), HA mAb (12CA5, Boehringer Mannheim), iNOS pAb (Zymed Laboratories) or nNOS mAb (Zymed Laboratories).

NO release from transfected COS cells. 24–48 h after transfection, media were processed for the measurement of nitrite (NO_2^-), the stable breakdown product of NO in aqueous solution, by NO-specific chemiluminescence as described²⁹. Media were deproteinized and samples containing NO_2^- were refluxed in glacial acetic acid containing sodium iodide. Under these conditions, NO_2^- was quantitatively reduced to NO which was quantified by a chemiluminescence detector after reaction with ozone in a NO analyser (Sievers). In all experiments, NO_2^- release was inhibitable by a NOS inhibitor. In addition, NO_2^- release from cells transfected with the β -galactosidase cDNA was subtracted to control for background levels of NO_2^- found in serum or media. In some experiments, we used cGMP accumulation in COS cells as a bioassay for the production of NO as described.

NOS activity assays. We used the conversion of ³H-L-arginine to ³H-L-citrulline to determine NOS activity in COS cell or endothelial cell lysates as described²⁶.

Phosphorylation studies *in vivo* and *in vitro*. For *in vivo* phosphorylation studies, COS cells were transfected with the cDNAs for wild-type or S635/

1179A eNOS and HA-Akt overnight. 36 h after transfection, we placed cells into dialysed serum-replete, phosphate-free Dulbecco's minimum essential medium supplemented with 80 μ Ci per ml of 32 P orthophosphoric acid for 3 h. Some cells were pretreated with wortmannin (500 nM) in the phosphate-free medium for 1 h and during the labelling. The lysates were collected and the eNOS was solubilized and partially purified by ADP Sepharose-affinity chromatography as previously described. 32 P incorporation into eNOS was visualized after SDS-PAGE (7.5%) by autoradiography, and the amount of eNOS protein was verified by western blotting for eNOS. For *in vitro* phosphorylation studies, we incubated recombinant eNOS purified from *Escherichia Coli* with wild-type or kinase-inactive Akt immunoprecipitated from transfected COS cells. eNOS was incubated with 32 P γ -ATP (2 μ l, specific activity 3000Ci per mmol), ATP (50 μ M) and DTT (1 mM), in a buffer containing HEPES (20 mM, pH = 7.4), MnCl₂ (10 mM), MgCl₂ (10 mM) and immunoprecipitated Akt for 20 min at room temperature. In experiments examining the *in vitro* phosphorylation of wild-type and mutant eNOS we incubated recombinant Akt (1 μ g) purified from baculovirus-infected SF9 cells with wild-type or S1179A eNOS (2.4 μ g, purified from *E. coli*) using essentially the same conditions as above. Proteins were resolved by SDS-PAGE and 32 P incorporation and the amount of protein was determined by Coomassie staining as above. In studies identifying the labelled eNOS peptide, we incubated immunoprecipitated Akt with recombinant eNOS as above. The sample was run on SDS-PAGE and the eNOS band digested in gel, and the resultant tryptic fragments were purified by RP-HPLC. Peptide mass and 32 P incorporation were monitored and the prominent labelled peak was further analysed by mass spectrometry. In other experiments, peptides corresponding to the potential Akt phosphorylation site were synthesized, purified by HPLC and verified by mass spectrometry (W. M. Keck Biotechnology Resource Center, Yale University School of Medicine). The wild-type peptide was ¹¹⁷⁴RIRTSQSLQERHLRGAVPWA¹¹⁹⁴ and the mutant peptide was identical except that Ser 1179 was changed to alanine. *In vitro* kinase reactions were done essentially as described above by incubating peptides (25 μ g) with recombinant Akt (1 μ g). Reactions were then spotted onto phosphocellulose filters and the amount of phosphate incorporated was measured by Cerenkov counting.

Adenoviral infections and NO release in endothelial cells. We cultured BLMVEC in either 100 mm dishes (for basal NO release and NOS activity assays) or C6 well plates (for stimulated NO) as described¹⁸. BLMVEC were infected with 200 MOI of adenovirus containing the β -galactosidase²⁹, HA-tagged, inactive phosphorylation mutant Akt (AA-Akt³⁰) or carboxy-terminal HA-tagged constitutively active Akt (myr-Akt) for 4 h. The virus was removed and cells left to recover for 18 h in complete medium. In preliminary experiments with the β -galactosidase virus, these conditions were optimal for infecting 100% of the cultures. For measurement of basal NO production, medium was collected 24 h after the initial infection with virus. For measurement of stimulated NO release, cells were then washed with serum-free medium followed by stimulation with VEGF (40 ng ml⁻¹) for 30 min. In some experiments, we determined the calcium dependency of NOS 24 h after adenoviral infection. Infected cells were lysed in NOS assay buffer containing 1% NP40, and detergent soluble material was used for activity assays. Lysates were incubated with EGTA-buffered calcium to yield appropriate amounts of free calcium in the incubation.

Statistics. Data are expressed as mean $\pm$ s.e.m. Comparisons were made using a two-tailed Student's *t*-test or ANOVA with a post-hoc test where appropriate. Differences were considered to be significant at *P* < 0.05.

Received 23 February; accepted 19 April 1999.

- Shesely, E. G. *et al.* Elevated blood pressures in mice lacking endothelial nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **93**, 13176–13181 (1996).
- Huang, P. L. *et al.* Hypertension in mice lacking the gene for endothelial nitric oxide synthase. *Nature* **377**, 239–242 (1995).
- Rudic, R. D. *et al.* Direct evidence for the importance of endothelium-derived nitric oxide in vascular remodeling. *J. Clin. Invest.* **101**, 731–736 (1998).
- Murohara, T. *et al.* Nitric oxide synthase modulates angiogenesis in response to tissue ischemia. *J. Clin. Invest.* **101**, 2567–2578 (1998).
- Michel, T., Li, G. K. & Busconi, L. Phosphorylation and subcellular translocation of endothelial nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **90**, 6252–6256 (1993).
- Garcia-Cardena, G., Fan, R., Stern, D. F., Liu, J. & Sessa, W. C. Endothelial nitric oxide synthase is regulated by tyrosine phosphorylation and interacts with caveolin-1. *J. Biol. Chem.* **271**, 27237–27240 (1996).
- Corson, M. A. *et al.* Phosphorylation of endothelial nitric oxide synthase in response to fluid shear stress. *Circ. Res.* **79**, 984–991 (1996).
- Coffer, P. J., Jin, J. & Woodgett, J. R. Protein kinase B (c-Akt): a multifunctional mediator of phosphatidylinositol 3-kinase activation. *Biochem. J.* **335**, 1–13 (1998).

- Frankel, T. F., Kaplan, D. R. & Cantley, L. C. PI3K: downstream AKTion blocks apoptosis. *Cell* **88**, 435–437 (1997).
- Downward, J. Mechanisms and consequences of activation of protein kinase B/Akt. *Curr. Opin. Cell Biol.* **10**, 262–267 (1998).
- Murga, C., Laguinge, L., Wetzker, R., Cuadrado, A. & Gutkind, J. S. Activation of Akt/protein kinase B by G protein-coupled receptors. A role for alpha and beta gamma subunits of heterotrimeric G proteins acting through phosphatidylinositol-3-OH kinase gamma. *J. Biol. Chem.* **273**, 19080–19085 (1998).
- Dimmeler, S., Assmus, B., Hermann, C., Haendeler, J. & Zeiher, A. M. Fluid shear stress stimulates phosphorylation of Akt in human endothelial cells: involvement in suppression of apoptosis. *Circ. Res.* **83**, 334–341 (1998).
- Cardone, M. H. *et al.* Regulation of cell death protease caspase-9 by phosphorylation. *Science* **282**, 1318–1321 (1998).
- Zeng, G. & Quon, M. J. Insulin-stimulated production of nitric oxide is inhibited by wortmannin. Direct measurement in vascular endothelial cells. *J. Clin. Invest.* **98**, 894–898 (1996).
- Papapetropoulos, A., Garcia-Cardena, G., Madri, J. A. & Sessa, W. C. Nitric oxide production contributes to the angiogenic properties of vascular endothelial growth factor in human endothelial cells. *J. Clin. Invest.* **100**, 3131–3139 (1997).
- Parenti, A. *et al.* Nitric oxide is an upstream signal of vascular endothelial growth factor-induced extracellular signal-regulated kinase1/2 activation in postcapillary endothelium. *J. Biol. Chem.* **273**, 4220–4226 (1998).
- Liu, J., Hughes, T. E. & Sessa, W. C. The first 35 amino acids and fatty acylation sites determine the molecular targeting of endothelial nitric oxide synthase into the golgi region of cells: A green fluorescent protein study. *J. Cell Biol.* **137**, 1525–1535 (1997).
- Garcia-Cardena, G., Oh, P., Liu, J., Schnitzer, J. E. & Sessa, W. C. Targeting of nitric oxide synthase to endothelial cell caveolae via palmitoylation: Implications for nitric oxide signaling. *Proc. Natl Acad. Sci. USA* **93**, 6448–6453 (1996).
- Shaul, P. W. *et al.* Acylation targets endothelial nitric-oxide synthase to plasmalemmal caveolae. *J. Biol. Chem.* **271**, 6518–6522 (1996).
- Sessa, W. C. *et al.* The Golgi association of endothelial nitric oxide synthase is necessary for the efficient synthesis of nitric oxide. *J. Biol. Chem.* **270**, 17641–17644 (1995).
- Liu, J., Garcia-Cardena, G. & Sessa, W. C. Palmitoylation of endothelial nitric oxide synthase is necessary for optimal stimulated release of nitric oxide: implications for caveolae localization. *Biochemistry* **35**, 13277–13281 (1996).
- Kantor, D. B. *et al.* A role for endothelial NO synthase in LTP revealed by adenovirus-mediated inhibition and rescue. *Science* **274**, 1744–1748 (1996).
- Gerber, H. P. *et al.* Vascular endothelial growth factor regulates endothelial cell survival through the phosphatidylinositol 3'-kinase/Akt signal transduction pathway. Requirement for flk-1/kdr activation. *J. Biol. Chem.* **273**, 30336–30343 (1998).
- Kuchan, M. J. & Frangos, J. A. Role of calcium and calmodulin in flow-induced nitric oxide production in endothelial cells. *Am. J. Physiol.* **266**, C628–C636 (1994).
- Ayajiki, K., Kindermann, M., Hecker, M., Fleming, I. & Busse, R. Intracellular pH and tyrosine phosphorylation but not calcium determine shear stress-induced nitric oxide production in native endothelial cells. *Circ. Res.* **78**, 750–758 (1996).
- Garcia-Cardena, G. *et al.* Dynamic activation of endothelial nitric oxide synthase by Hsp90. *Nature* **392**, 821–824 (1998).
- Yano, S., Tokumitsu, H. & Soderling, T. R. Calcium promotes cell survival through CaM-K kinase activation of the protein kinase-B pathway. *Nature* **396**, 584–587 (1998).
- Salerno, J. C. *et al.* An autoinhibitory control element defines calcium-regulated isoforms of nitric oxide synthase. *J. Biol. Chem.* **272**, 29769–29777 (1997).
- Smith, R. C. *et al.* p21CIP1-mediated inhibition of cell proliferation by overexpression of the gap homeodomain gene. *Genes Dev.* **11**, 1674–1689 (1997).
- Alessi, D. R. *et al.* Mechanism of activation of protein kinase B by insulin and IGF-1. *EMBO J.* **15**, 6541–6551 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank P. Martasek and B. S. Masters for recombinant eNOS; J. Liu for constructing and characterizing the myr-nNOS construct; D. Bredt and T. Billiar for NOS cDNAs; T. Zioncheck for human VEGF; Y. Chen for generating the eNOS S1179A *E. Coli* expression construct; and K. Williams, M. LoPrete and K. Stone in the Keck Facility for their help with identification of the eNOS phosphopeptides. This work was supported by grants from the NIH and American Heart Association. W.C.S. is an Established Investigator of the American Heart Association. J.P.G. is supported by fellowships from the Heart and Stroke Foundation of Canada and from FCAR.

Correspondence and requests for materials should be addressed to W.C.S. (e-mail: william.sessa@yale.edu).

Activation of nitric oxide synthase in endothelial cells by Akt-dependent phosphorylation

Stefanie Dimmeler*, Ingrid Fleming†, Beate Fisslthaler†, Corinna Hermann*, Rudi Busse† & Andreas M. Zeiher*

* Molecular Cardiology, Department of Internal Medicine IV, and † Institute of Cardiovascular Physiology, University of Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

Nitric oxide (NO) produced by the endothelial NO synthase (eNOS) is a fundamental determinant of cardiovascular homeostasis: it regulates systemic blood pressure, vascular remodelling and angiogenesis^{1–3}. Physiologically, the most important stimulus for the continuous formation of NO is the viscous drag (shear

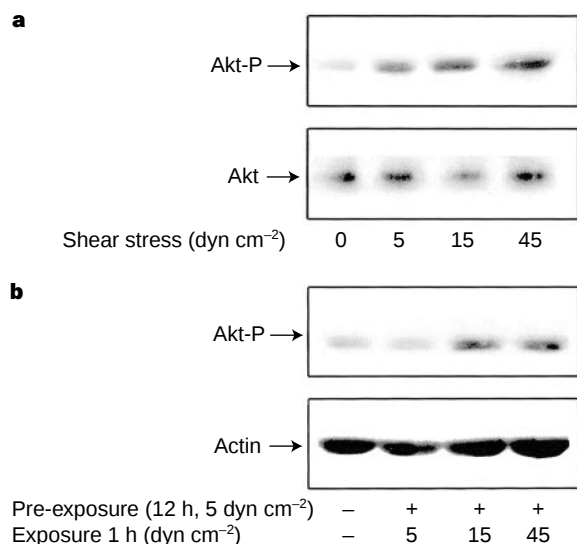


Figure 1 Shear stress stimulates the phosphorylation of Akt. **a**, Stimulus intensity-dependent increase in Akt phosphorylation. Cultured HUVEC were exposed to shear stress for 1 h and Akt phosphorylation was detected by western blot analysis with phospho-specific Akt antibodies. Blots were reprobed with Akt to confirm equal loading of the gel. **b**, Akt phosphorylation in shear-stress-preconditioned cells. HUVEC were pre-exposed to shear stress (5 dyn cm⁻²) for 12 h, followed by an increase to either 15 or 45 dyn cm⁻² for a further 1 h. Akt phosphorylation was detected by western blot analysis. The specificity of the phospho-specific Akt antibody was confirmed as described previously¹⁹.

stress) generated by the streaming blood on the endothelial layer⁴⁻⁸. Although shear-stress-mediated phosphorylation of eNOS is thought to regulate enzyme activity^{9,10}, the mechanism of activation of eNOS is not yet known. Here we demonstrate that the serine/threonine protein kinase Akt/PKB¹¹⁻¹³ mediates the activation of eNOS, leading to increased NO production. Inhibition of the phosphatidylinositol-3-OH kinase/Akt pathway or mutation of the Akt site on eNOS protein (at serine 1177) attenuates the serine phosphorylation and prevents the activation of eNOS. Mimicking the phosphorylation of Ser 1177 directly enhances enzyme activity and alters the sensitivity of the enzyme to Ca²⁺, rendering its activity maximal at sub-physiological concentrations of Ca²⁺. Thus, phosphorylation of eNOS by Akt represents a novel Ca²⁺-independent regulatory mechanism for activation of eNOS.

Endothelium-derived NO has a crucial role in the local regulation of vascular homeostasis and platelet aggregation¹. A decrease in the bioavailability of NO is a characteristic feature in patients with coronary artery disease¹⁴ and aggravates the development of atherosclerotic lesions¹⁵. Consistent with its classification as a Ca²⁺/calmodulin-dependent enzyme^{16,17}, eNOS can be activated by various agonists after an increase in the intracellular concentration of free Ca²⁺ ([Ca²⁺]_i). But the physiologically most important pathway for eNOS activation in response to fluid shear stress is different from that elicited by other stimuli in that it is maintained (over hours) and can be observed in the absence of extracellular Ca²⁺, and is not inhibited by calmodulin antagonists^{5,18}. The phosphorylation of eNOS is thought to regulate enzyme activity in a Ca²⁺-independent fashion^{9,10}, but the signalling pathway involved is not known. As phosphatidylinositol-3-OH kinase (PI(3)K) and the downstream serine/threonine kinase Akt/PKB are activated in endothelial cells in response to shear stress¹⁹, we investigated the involvement of the PI(3)K/Akt pathway in the regulation of eNOS activity.

Exposure of human umbilical vein endothelial cells (HUVEC) to shear stress in a cone-plate viscometer elicited the rapid phosphorylation and activation of Akt; the maximum effect was observed

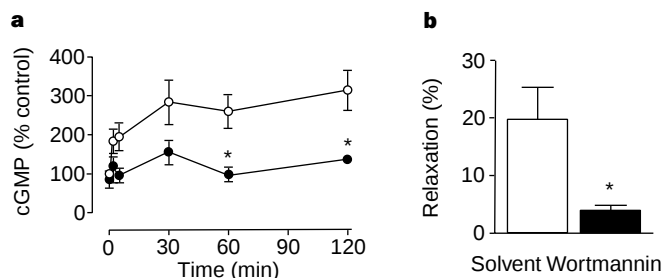


Figure 2 Inhibition of PI(3)K prevents shear-stress-induced cGMP increase and NO release. **a**, HUVEC were exposed to shear (15 dyn cm⁻²) in the presence (filled circles) or absence (open circles) of wortmannin (20 μM), and cGMP levels were measured as described¹⁰. IBMX (0.1 mM) was present only during the last 30 min. Data are mean ± s.e.m. (n = 5; asterisks, P < 0.05 compared with shear stress). **b**, Shear-stress-induced NO release in a bioassay system with endothelium-intact coronary artery segments as donor and assessed by the relaxation of a maximally precontracted coronary artery ring with the endothelium removed. Results are presented as mean ± s.e.m. (n = 7; asterisk, P < 0.05).

after 1–2 hours, as assessed by western blotting and a selective assay of enzyme activity (Fig. 1a)¹⁹. The phosphorylation of Akt was also dependent on the degree of shear stress (Fig. 1a). Moreover, a step increase in fluid shear stress from 5 to 15 dyn cm⁻², or from 5 to 45 dyn cm⁻² for 1 hour, also induced an increase in phosphorylation of Akt (Fig. 1b).

The PI(3)K-inhibitor wortmannin, which prevents the phosphorylation and activation of Akt¹⁹, completely inhibited the shear-stress-induced increase in cyclic GMP levels (Fig. 2a). To determine the effects of wortmannin on shear-stress-induced NO production in native endothelial cells, we used a bioassay system in which shear stress was increased by inducing vasoconstriction in an endothelium-intact donor segment at constant flow. In this model, wortmannin abolished the shear-stress-induced production of NO (Fig. 2b). Thus, the shear stress-induced activation of eNOS seems to be dependent on the activation of the PI(3)K/Akt pathway.

To examine whether the activation of Akt is sufficient to enhance eNOS activity, we overexpressed a constitutively active Akt construct in which Thr 308 and Ser 473 were replaced with aspartate²⁰. Overexpression of the constitutive active Akt resulted in a 1.4-fold increase in intracellular cGMP, whereas overexpression of a dominant-negative Akt mutant¹⁹ decrease basal cGMP levels by ~30%, indicating that eNOS can be activated by an Akt-dependent pathway. Vascular-endothelium-derived growth factor (VEGF), which activates both eNOS^{21,22} and the PI(3)K/Akt pathway²³, also elicited a PI(3)K-dependent increase in cGMP levels and enhanced both the phosphorylation and activity of Akt (results not shown).

To determine the role of shear-stress-induced activation of the PI(3)K/Akt pathway in the phosphorylation of eNOS, ³²P-labelled HUVEC were exposed to shear stress (12 dyn cm⁻²) for up to 1 hour. Shear stress increased the phosphorylation of eNOS about 2-fold (Fig. 3a). This increase was due to the phosphorylation of serine residues, as demonstrated by two-dimensional phosphoamino acid analysis (Fig. 3b). Under these experimental conditions, the eNOS recovered from HUVEC was not phosphorylated on either tyrosine or threonine residues. Incubation with wortmannin completely prevented shear-stress-induced eNOS phosphorylation (Fig. 3a). To assess the role of Akt in eNOS phosphorylation, both eNOS and the active Akt kinase were immunoprecipitated from endothelial cells. In the presence of [γ-³²P]ATP, active Akt directly phosphorylated eNOS and a second Akt substrate, histone 2B, *in vitro* (Fig. 4a, b), whereas the kinase-inactive mutants T308A/S473A or K179M did not affect eNOS phosphorylation (Fig. 4a). Moreover, ³²P-labelling experiments *in vivo* revealed that constitutively active Akt stimulates the phosphorylation of eNOS, but the inactive kinase does not (Fig. 4b).

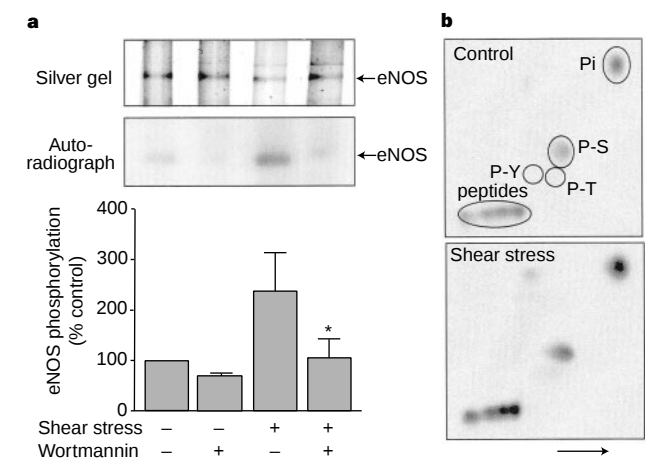


Figure 3 Inhibition of PI(3)K by wortmannin inhibits the shear-stress-induced phosphorylation of eNOS. **a**, HUVEC were labelled with ^{32}P and exposed to shear stress (15 dyn cm^{-2}) and wortmannin (20 nM) for 1 h, as indicated. eNOS protein was then immunoprecipitated, separated by SDS-PAGE and blotted onto a membrane. eNOS phosphorylation was detected by autoradiography and quantified by phosphorimaging. The incorporation of radioactivity was corrected for eNOS protein levels, which were determined on a silver-stained gel. Values are expressed as mean $\pm$ s.e.m. ($n = 4$; asterisk, $P < 0.05$ compared with shear stress). **b**, Detection of phosphorylated amino-acid residues. In some experiments, the eNOS protein band was excised and phosphorylated amino-acid residues were detected by two-dimensional gel electrophoresis. The autoradiogram shown is representative of four experiments.

To test eNOS further as a potential substrate for Akt phosphorylation, we specifically mutated the two putative Akt-kinase sites on eNOS at residues Ser 633 and Ser 177, corresponding to the consensus sequence for Akt sites (RXRXS/T), to aspartate (S633D, S1177D) to mimic the continuous phosphorylation of the protein. eNOS activity was similar in HUVEC transfected with either the wild-type enzyme or the eNOS construct in which Ser 633 was replaced by aspartate (Fig. 4c). In contrast, simulation of phosphorylation on Ser 1177 (S1177D) significantly augmented enzyme activity (Fig. 4c). Overexpression of a construct in which Ser 1177 phosphorylation was prevented by replacement with alanine resulted in an enzyme activity similar to that of the wild type (Fig. 4c). The differences in enzymatic activity of the overexpressed eNOS mutants were not due to different expression levels, as indicated by western blotting (S633D $95 \pm 12\%$, S1177D $92 \pm 13\%$, S1177A $93 \pm 23\%$, compared with wild-type eNOS) (Fig. 4d).

Mutation of Ser 1177 decreased the Akt-dependent phosphorylation of the eNOS *in vitro* (Fig. 4e). Furthermore, mutation of Ser 1177 prevented the Akt-induced and shear-stress-induced serine phosphorylation of eNOS (Fig. 5a, b, respectively). These results were confirmed in ^{32}P -labelled HUVEC; exposure of HUVEC to shear induced a 1.64-fold increase in phosphorylation of the eNOS wild-type protein, but not the S1177A mutant (49% compared with static conditions). Thus, Ser 1177 represents the primary phosphorylation site of eNOS in intact cells. Finally, Akt-stimulated nitrite accumulation in COS cells transfected with wild-type eNOS was abolished by mutating Ser 1177 (Fig. 5c), emphasizing that Ser 1177 is an acceptor amino acid for Akt-mediated phosphorylation with a distinct consequence for eNOS activity. In contrast, mutating Ser 633 did not significantly affect Akt-mediated NOS phosphorylation *in vitro* ($111 \pm 11\%$, compared with wild-type eNOS; Fig. 4e) and did not decrease Akt-stimulated eNOS activity (Fig. 5c). To demonstrate further that endogenous Akt is capable of phosphorylating eNOS, eNOS-transfected COS cells were stimulated with serum, a potent activator of Akt. As shown in Fig. 5d, serum-induced phosphorylation of eNOS was significantly

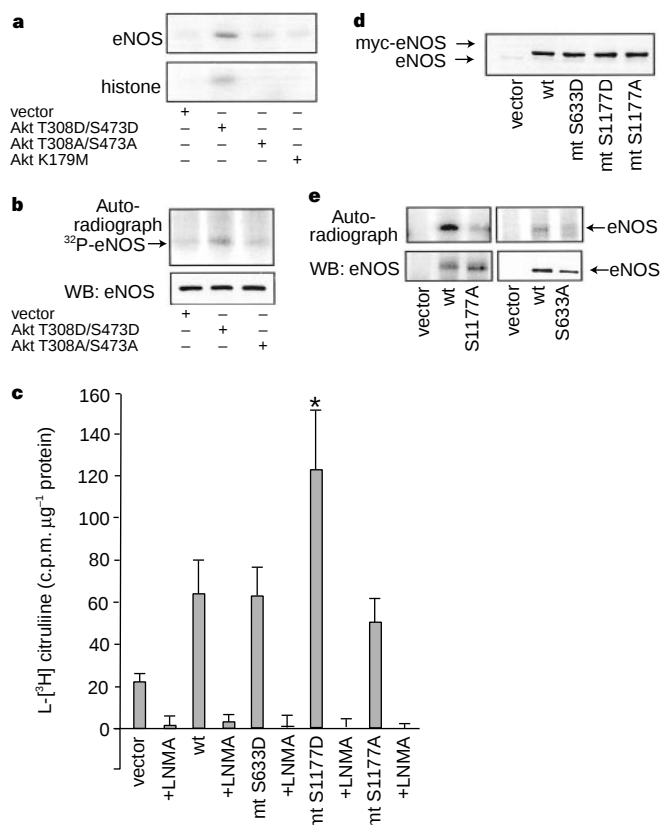


Figure 4 Akt phosphorylates Ser 1177 and increases eNOS enzyme activity. **a**, Phosphorylation of eNOS by Akt. Constitutively active Myc-tagged Akt kinase (T308D/S473D) or inactive mutants (T308A/S473A or K179M) were immunoprecipitated from transfected COS cells and incubated with immunoprecipitated eNOS (upper panel) or histone 2B (lower panel), in the presence of [γ - ^{32}P]ATP. Representative autoradiograms are shown. Western blot analysis confirmed equal expression of the construct (results not shown). **b**, COS cells were co-transfected with wild-type eNOS and constitutively active (T308D/S473D) or inactive (T308A/S473A) Akt. After ^{32}P -labelling, eNOS was immunoprecipitated with anti-eNOS antibody. Western blot analysis confirmed equal loading. **c**, Wild-type (wt) eNOS or eNOS mutants (mt) were expressed in HUVEC, protein was isolated and eNOS activity was measured with a citrulline assay in the presence or absence of N^G -monomethyl-L-arginine (LNMA, $200 \mu\text{M}$). Data are mean $\pm$ s.d. (asterisk, $P < 0.05$ compared with wild type). **d**, Western blot analysis for eNOS to control for expression of the different Myc-tagged wild-type eNOS or eNOS mutant and endogenous eNOS (eNOS). **e**, Phosphorylation of eNOS by Akt. COS cells were transfected with vector (pcDNA3.1-myc-his), wild-type eNOS or eNOS mutants. Myc-tagged proteins were immunoprecipitated and served as substrates. Active Akt kinase was immunoprecipitated from COS cells stimulated with fetal calf serum with phospho-specific Akt antibodies; Akt immunoprecipitates were added to the respective substrates and incubated with [γ - ^{32}P]ATP to allow protein phosphorylation. Representative autoradiograms are shown. Western blot analysis (WB) of an aliquot of the immunoprecipitates with anti-eNOS antibodies confirmed equal loading (lower panel).

decreased by the PI(3)K inhibitor Ly294002 or by co-transfection of a dominant-negative Akt mutant. In addition, shear-stress-stimulated phosphorylation of eNOS was decreased by overexpression of a dominant-negative Akt mutant (results not shown). Taken together, these results indicate that eNOS is indeed a target of Akt. However, although it is unlikely, we cannot formally exclude the possibility that, *in vivo*, a kinase activated by Akt with Akt-like specificity acts between Akt and eNOS.

Having shown that phosphorylation of the Ser 1177 residue increases the activity of eNOS, we investigated the underlying mechanism. As shear stress stimulates NO synthesis independently of an increase in $[\text{Ca}^{2+}]_i$, we considered that the Akt-dependent

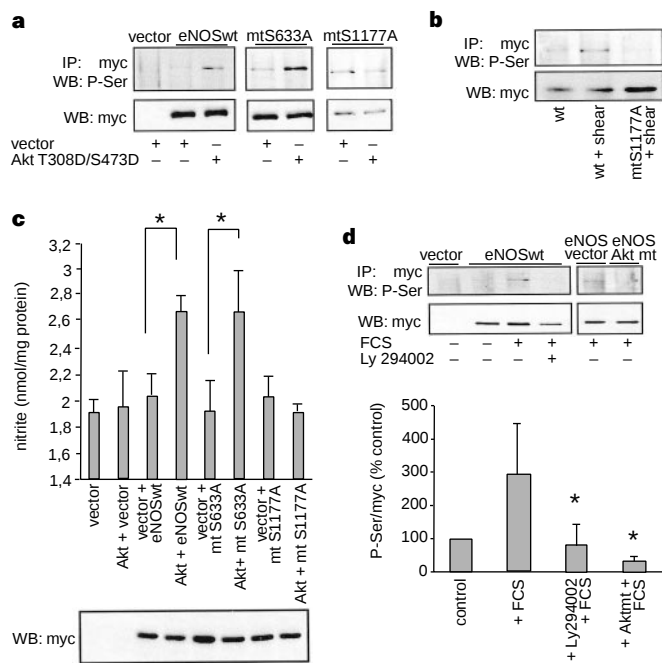


Figure 5 Akt phosphorylates eNOS and increases eNOS enzyme activity in cells. **a**, COS cells were co-transfected with active Akt (T308D/S473D) or vector and the eNOS constructs as indicated. Myc-tagged eNOS was immunoprecipitated (IP) and serine phosphorylation was detected by western blot (WB) analysis with anti-phosphoserine antibodies (clone 4H4; Biomol, Hamburg, Germany). Blots were reprobed with anti-Myc antibodies (lower panel). **b**, HUVEC were transfected with the respective plasmids and exposed to shear stress for 1 h. Myc-tagged wild-type (wt) eNOS and eNOS mutants (mt) were immunoprecipitated, followed by western blotting with anti-phosphoserine antibodies. Blots were reprobed with anti-Myc antibody as loading control (lower panel). **c**, COS cells were co-transfected with constitutively active Akt (T308D/S473D) and the respective eNOS constructs. Nitrite accumulation over 24 h was measured with the Griess reaction. The expression of the eNOS constructs was controlled for by western blotting with anti-Myc antibodies (lower panel). Data are mean $\pm$ s.e.m. ($n = 4$; asterisks, $P < 0.05$). **d**, COS cells were transfected with vector or eNOS and starved for 4 h. After stimulation of COS cells with 10% v/v fetal calf serum (FCS) for 45 min in the presence or absence of Ly294002 (10 μ M), eNOS phosphorylation was detected in Myc immunoprecipitates with the anti-phospho-serine antibody (see **a**). The effect of the dominant-negative Akt mutant (Aktmt) was tested under similar conditions by using co-transfected COS cells (eNOS, 4 μ g; vector or Akt mt, 6 μ g DNA). The immunoprecipitation was controlled by reprobing the blots with anti-Myc antibodies (lower panel). The blots were analysed densitometrically and the ratio of phosphorylated eNOS to Myc-tagged eNOS protein was determined to normalize for differences in loading (bar graph). Data are mean $\pm$ s.e.m. ($n = 5$; asterisks, $P < 0.01$ compared with in the presence of FCS).

phosphorylation of eNOS might be responsible for its Ca^{2+} -independent activation. As shown in Fig. 5a, the activity of wild-type eNOS was decreased by a stepwise decrease in the Ca^{2+} concentration in the assay buffer. This effect was not observed with the Ser 1177D mutant (Fig. 6a). Complete chelation of Ca^{2+} with EDTA abrogated activity of wild-type and Ser 1177 eNOS (Fig. 6a). Similar results have been presented for the Ca^{2+} -independent inducible NOS, which does not require an increase in $[\text{Ca}^{2+}]_i$ to optimize NO production, although low concentrations of Ca^{2+} are essential for enzyme activity²⁴. These results suggest that the shear-stress-induced Akt-dependent phosphorylation of Ser 1177 enhances eNOS enzyme activity independently of an increase in $[\text{Ca}^{2+}]_i$. Moreover, competitive antagonism of calmodulin binding by calmidazolium decreased the activity of the wild-type protein more effectively than that of the Ser 1177 mutant ($65 \pm 17\%$ inhibition, compared with $39 \pm 5\%$ inhibition by 20 μ M calmidazolium, respectively; $P < 0.05$).

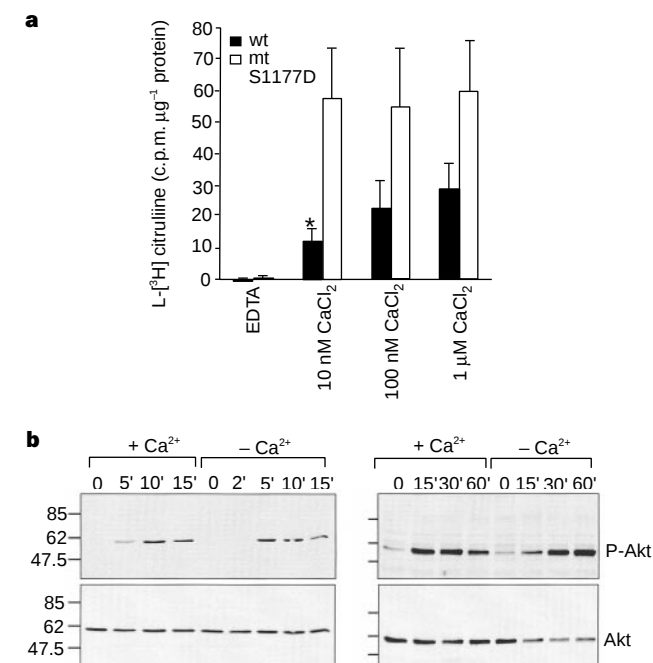


Figure 6 Calcium dependence of eNOS activity and Akt phosphorylation. **a**, HUVEC were transfected with wild-type (wt) eNOS or eNOS mutant (mt S1177D), and eNOS activity was measured in protein homogenates in the presence of the indicated concentrations of CaCl_2 . The lower basal activity of the homogenates compared with those in Fig. 4b is due to differences in the preparation (see Method). Data are mean $\pm$ s.e.m. ($n = 4$; asterisk, $P < 0.05$ compared with 1 μ M CaCl_2). **b**, Shear-stress-stimulated Akt phosphorylation was measured in the presence or absence of CaCl_2 and was assessed with the phospho-specific Akt antibody.

zium, respectively; $P < 0.05$). Akt has been reported to be activated by Ca^{2+} -dependent stimulation of calmodulin-dependent protein kinase kinase²⁵. However, the shear-stress-induced phosphorylation of Akt was completely unaffected by the removal of extracellular Ca^{2+} (Fig. 6b) or by the inclusion of calmidazolium (results not shown), precluding a role for Ca^{2+} and calmodulin in the signalling pathway activated by fluid shear stress and leading to the activation of eNOS.

Taken together, our results show that the stimulation of PI(3)K/Akt by shear stress and VEGF elicits the serine phosphorylation of eNOS, and thereby enhances enzyme activity in a Ca^{2+} -independent manner. As only a few substrates for the protein kinase Akt/PKB have been described so far, the results reported here identify eNOS as a novel Akt target. Given that the activation of this pathway results in the maintained production of NO, it is most likely to be implicated in the regulation of NO-controlled gene expression²⁶ as well as the NO-mediated angiogenic response to VEGF^{23,21,22}. Moreover, as NO is necessary for vascular remodelling and its decreased bioavailability facilitates the pathophysiological changes in vascular wall morphology associated with hypertension and atherosclerosis², the Akt-mediated increase in eNOS activity might represent a major determinant of vessel architecture. Finally, the activation of eNOS by Akt might contribute to the important physiological effects of Akt on apoptosis, cell attachment and cell proliferation^{27–29}.

Methods

Cell culture and western blot analysis. HUVEC were purchased from Cell systems/Clonetics, Solingen, Germany, and were cultured as described previously²⁸. Similar results were obtained when HUVEC were isolated and cultured in M199 medium. Fluid shear stress was applied by using a cone-plate viscometer as described previously¹⁰. For western blotting, HUVEC homogenates

were obtained as described²⁸ and blots were incubated with phospho-specific anti-Akt antibody (Biolabs, Schwalbach, Germany), anti-Akt antibody (Biolabs) or anti-eNOS antibody (Dianova, Hamburg, Germany).

Biossay. Pig epicardial artery segments (~40 mm in length; external diameter 2.4–2.8 mm) were cannulated at both ends and placed into organ chambers as described¹⁸. After equilibration, the luminal effluents from this donor segment directly superfused an endothelium-denuded detector ring (delay ~2 s). Release of NO from the endothelium of the donor segment was elicited either by vasoconstriction to increase shear stress at constant flow or by the application of bradykinin as a bolus. NO-mediated relaxant responses of the detector ring to the effluents from the donor segment were standardized by comparison with the response observed after the application of glyceryl trinitrate (10 and 100 pmol) directly to the detector tissue.

Plasmid transfection. Plasmid encoding the human eNOS was a gift from Dr Nakane. After restriction with *Hind*III and *Eco*RV, eNOS was subcloned into the respective sites of pcDNA3.1-Myc-His (Invitrogen). The putative Akt sites were changed by site-directed mutagenesis (Stratagene, Heidelberg, Germany). The plasmids encoding the bovine Akt and the truncated dominant-negative Akt mutant were donated by J. Downward²⁷ and were subcloned in pcDNA3.1 vector as described¹⁹. Clones with verified sequences were transfected in HUVEC (3.5×10^5 cells per 6 cm well; 3 µg plasmid DNA; 20 µl Superfect) or COS cells (3.8×10^5 cells per 6 cm, 8 µg plasmid DNA, 30 µl Superfect) as described previously¹⁹.

L-Arginine assay. Transfected HUVEC were labelled in homogenization buffer (50 mM Tris, pH 7.4, 1.15% w/v KCl, 5 mM glucose, 1% v/v Triton X-100, 1 mM PMSE, 2 mM dithiothreitol) with (Fig. 4) or without (Fig. 6) 1 mM EDTA. NOS activity in protein homogenates (100 µg, 100 µl) was assessed by detecting the conversion of L-[³H]arginine to L-citrulline in the presence of cofactors (in 50 µl, 1 µM calmodulin, 1 µM FAD, 1 mM NADPH, 15 µM tetrahydrobiopterin; in Fig. 4c, 1.25 mM CaCl₂).

³²P-labelling, eNOS immunoprecipitation and phosphoamino acid analysis. HUVEC or COS cells were loaded with ³²P (H₃³²PO₄, 30 µM) for 6 h in phosphate-free medium as described¹⁰. After stimulation, cells were harvested and lysed in ice-cold homogenization buffer (Tris-HCl, 50 mM, pH 7.5; 150 mM NaCl, 2 mM EDTA, 8 mM EGTA, 1% v/v Triton X-100, 25 mM NaF, 10 mM Na₂P₂O₇, 1 mM Na₃VO₄, 2 µg ml⁻¹ leupeptin, 2 µg ml⁻¹ pepstatin A, 10 µg ml⁻¹ trypsin inhibitor, and 44 µg ml⁻¹ PMSF). Endogenous eNOS protein or Myc-tagged overexpressed eNOS was immunoprecipitated and separated by SDS-PAGE. For two-dimensional gel electrophoresis, labelled eNOS protein was transferred to PVDF membranes (Millipore, Eschborn, Germany) and hydrolysed in 6 M HCl. Samples were resuspended in 2.2% v/v formic acid and 7.5% v/v acetic acid, spiked with marker amino acids and subjected to two-dimensional thin-layer electrophoresis at pH 1.9 and then at pH 3.5 with thin-layer cellulose plates (Merck, Darmstadt, Germany) as described³⁰. Marker amino acids were stained with ninhydrin and radioactivity was detected and quantified with a PhosphorImager (Fuji, BAS 1500). For detection of phosphorylation of eNOS *in vitro* by Akt, COS cells were transfected with Myc-tagged wild-type or mutant eNOS and proteins (1 mg per sample) were immunoprecipitated with anti-Myc antibodies. Active Akt kinase was obtained by immunoprecipitation of phosphorylated Akt with phospho-specific antibodies¹⁹ or by immunoprecipitation of overexpressed Myc-tagged constitutively active Akt (T308D/S473D). The immunoprecipitates were combined and incubated for 30 min at 30 °C in 30 µl of kinase reaction mixture containing 25 mM Tris, pH 7.5, 5 mM β-glycerophosphate, 0.1 mM Na₃VO₄, 2 mM dithiothreitol, 10 mM MgCl₂, 10 mM MnCl₂, 50 µM ATP and 5 µCi of [γ-³²P]ATP. The reaction was terminated by the addition of SDS sample buffer and samples were subjected to 8% SDS-PAGE and analysed by phosphorimaging.

Received 23 December 1998; accepted 6 May 1999.

- Moncada, S. & Higgs, A. The L-arginine-nitric oxide pathway. *N. Engl. J. Med.* **329**, 2002–2012 (1993).
- Rudic, R. D. *et al.* Direct evidence for the importance of endothelium-derived nitric oxide in vascular remodeling. *J. Clin. Invest.* **101**, 731–736 (1998).
- Murohara, T. *et al.* Nitric oxide synthase modulates angiogenesis in response to tissue ischemia. *J. Clin. Invest.* **101**, 2567–2578 (1998).
- Rubanyi, G. M., Romero, J. C. & Vanhoutte, P. M. Flow-induced release of endothelium-derived relaxing factor. *Am. J. Physiol.* **250**, H1145–H1149 (1986).
- Kuchan, M. J. & Frangos, J. A. Role of calcium and calmodulin in flow-induced nitric oxide production in endothelial cells. *Am. J. Physiol.* **266**, C628–C636 (1994).
- Sessa, W. C., Pritchard, K., Seyedi, N., Wang, J. & Hintze, T. H. Chronic exercise in dogs increases

coronary vascular nitric oxide production and endothelial cell nitric oxide synthase gene expression. *Circ. Res.* **74**, 349–353 (1994).

- Davies, P. F. Flow-mediated endothelial mechanotransduction. *Physiol. Rev.* **75**, 519–560 (1995).
- Busse, R. & Fleming, I. Pulsatile stretch and shear: physiological stimuli determining the production of endothelial-derived relaxing factor. *J. Vasc. Res.* **35**, 73–84 (1998).
- Corson, M. A. *et al.* Phosphorylation of endothelial nitric oxide synthase in response to fluid shear stress. *Circ. Res.* **79**, 984–991 (1996).
- Fleming, I., Bauersachs, J., Fisslthaler, B. & Busse, R. Ca²⁺-independent activation of the endothelial nitric oxide synthase in response to tyrosine phosphatase inhibitors and fluid shear stress. *Circ. Res.* **79**, 984–991 (1996).
- Frank, T. F. *et al.* The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell* **81**, 727–736 (1995).
- Burgering, B. M. T. & Coffey, P. J. Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature* **376**, 599–602 (1995).
- Downward, J. Lipid-regulating kinases: some common themes at last. *Science* **279**, 673–674 (1998).
- Zeher, A. M. Endothelial vasodilator dysfunction: pathogenetic link to myocardial ischaemia or epiphenomenon? *Lancet* **348**, S10–S12 (1996).
- Moroi, M. *et al.* Interaction of genetic deficiency of endothelial nitric oxide, gender, and pregnancy in vascular responses to injury in mice. *J. Clin. Invest.* **101**, 1225–1232 (1998).
- Bredt, D. S. *et al.* Cloned and expressed nitric oxide synthase structurally resembles cytochrome P-450 reductase. *Nature* **351**, 714–718 (1991).
- Busse, R. & Mülsch, A. Calcium-dependent nitric oxide synthesis in endothelial cytosol is mediated by calmodulin. *FEBS Lett.* **265**, 133–136 (1990).
- Ayajiki, K., Kindermann, M., Hecker, M., Fleming, I. & Busse, R. Intracellular pH and tyrosine phosphorylation but not calcium determine shear stress-induced nitric oxide production in native endothelial cells. *Circ. Res.* **78**, 750–758 (1996).
- Dimmeler, S., Assmus, B., Hermann, C., Haendeler, J. & Zeher, A. M. Fluid shear stress stimulates phosphorylation of Akt in human endothelial cells: involvement in suppression of apoptosis. *Circ. Res.* **83**, 334–342 (1998).
- Alessi, D. R. *et al.* Mechanism of activation of protein kinase B by insulin and IGF-1. *EMBO J.* **15**, 6541–6551 (1996).
- van der Zee, R. *et al.* Vascular endothelial growth factor/vascular permeability factor augments nitric oxide release from quiescent rabbit and human vascular endothelium. *Circulation* **95**, 1030–1037 (1997).
- Ziche, M. *et al.* Nitric oxide synthase lies downstream from vascular endothelial growth factor-induced but not basic fibroblast growth factor-induced angiogenesis. *J. Clin. Invest.* **99**, 2625–2634 (1997).
- Gerber, H. P., Dixit, V. & Ferrara, N. Vascular endothelial growth factor induces expression of the antiapoptotic proteins Bcl-2 and A1 in vascular endothelial cells. *J. Biol. Chem.* **273**, 13313–13316 (1998).
- Förstermann, U. *et al.* Isoforms of nitric oxide synthase. Characterization and purification from different cell types. *Biochem. Pharmacol.* **42**, 1849–1857 (1991).
- Yano, S., Tokumitsu, H. & Soderling, T. R. Calcium promotes cell survival through CaM-K kinase activation of the protein kinase-B pathway. *Nature* **396**, 584–587 (1998).
- Zeher, A. M., Fisslthaler, B., Schray-Utz, B. & Busse, R. Nitric oxide modulates the expression of monocyte chemoattractant protein 1 in cultured human endothelial cells. *Circ. Res.* **76**, 980–986 (1995).
- Khawaja, A., Rodriguez-Viciana, P., Wennström, S., Warne, P. H. & Downward, J. Matrix adhesion and Ras transformation both activate a phosphoinositide 3-OH kinase and protein kinase B/Akt cellular survival pathway. *EMBO J.* **16**, 2783–2793 (1997).
- Dimmeler, S., Haendeler, J., Nehls, M. & Zeher, A. M. Suppression of apoptosis by nitric oxide via inhibition of ICE-like and CPP32-like proteases. *J. Exp. Med.* **185**, 601–608 (1997).
- Kauffmann-Zeh, A. *et al.* Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB. *Nature* **385**, 544–548 (1997).
- Copper, J. A., Sefton, B. M. & Hunter, T. Detection and quantification of phosphotyrosine in proteins. *Methods Enzymol.* **99**, 387–402 (1983).

Acknowledgements. We thank C. Goebel, I. Winter and S. Ficus for expert technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to A.M.Z. (e-mail: Zeher@em.uni-frankfurt.de).

Enhancement of TBP binding by activators and general transcription factors

Xiao-Yong Li, Amy Virbasius, Xiaochuan Zhu & Michael R. Green

Howard Hughes Medical Institute, Program in Molecular Medicine, University of Massachusetts Medical Center, 373 Plantation Street, Suite 309, Worcester, Massachusetts 01605, USA

Eukaryotic transcriptional activators function, at least in part, by promoting assembly of the preinitiation complex^{1–3}, which comprises RNA polymerase II and its general transcription factors (GTFs). Activator-mediated stimulation of the assembly of the preinitiation complex has been studied *in vitro* but has been relatively refractory to *in vivo* analysis. Here we use a DNA-crosslinking/immunoprecipitation assay to study in living cells the first step in the assembly of the preinitiation complex, the

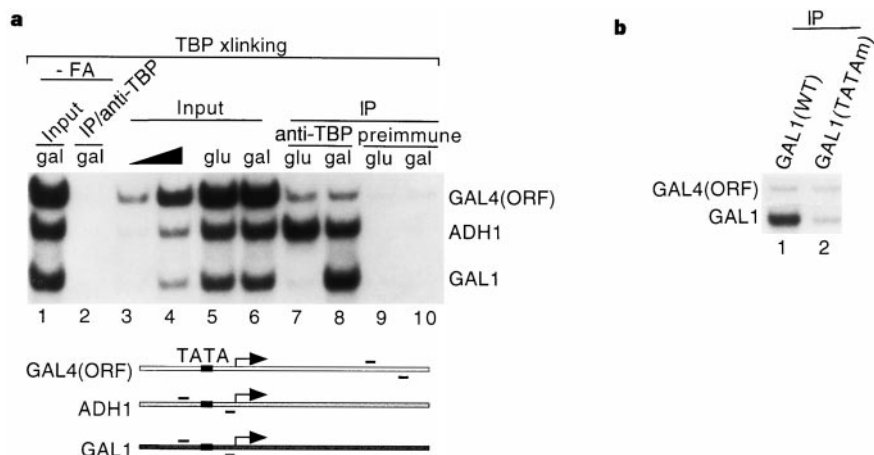


Figure 1 An *in vivo* crosslinking assay for measuring TBP-TATA-box interaction. **a**, *GAL1*, *ADH1*. The locations of the primers used for PCR amplification of the immunopurified DNA fragments are schematically shown. FA, formaldehyde. **b**, *GAL1*(TATAm). The crosslinking analysis was done using cells carrying a wild-

type *GAL1* promoter (*GAL1*(WT)) or a *GAL1* promoter mutant containing a 3-bp substitution in the TATA-box sequence (*GAL1*(TATAm)) fused to the *lacZ* gene⁴. Cells were grown in YP plus 2% glycerol, 2% galactose and 2% acetate.

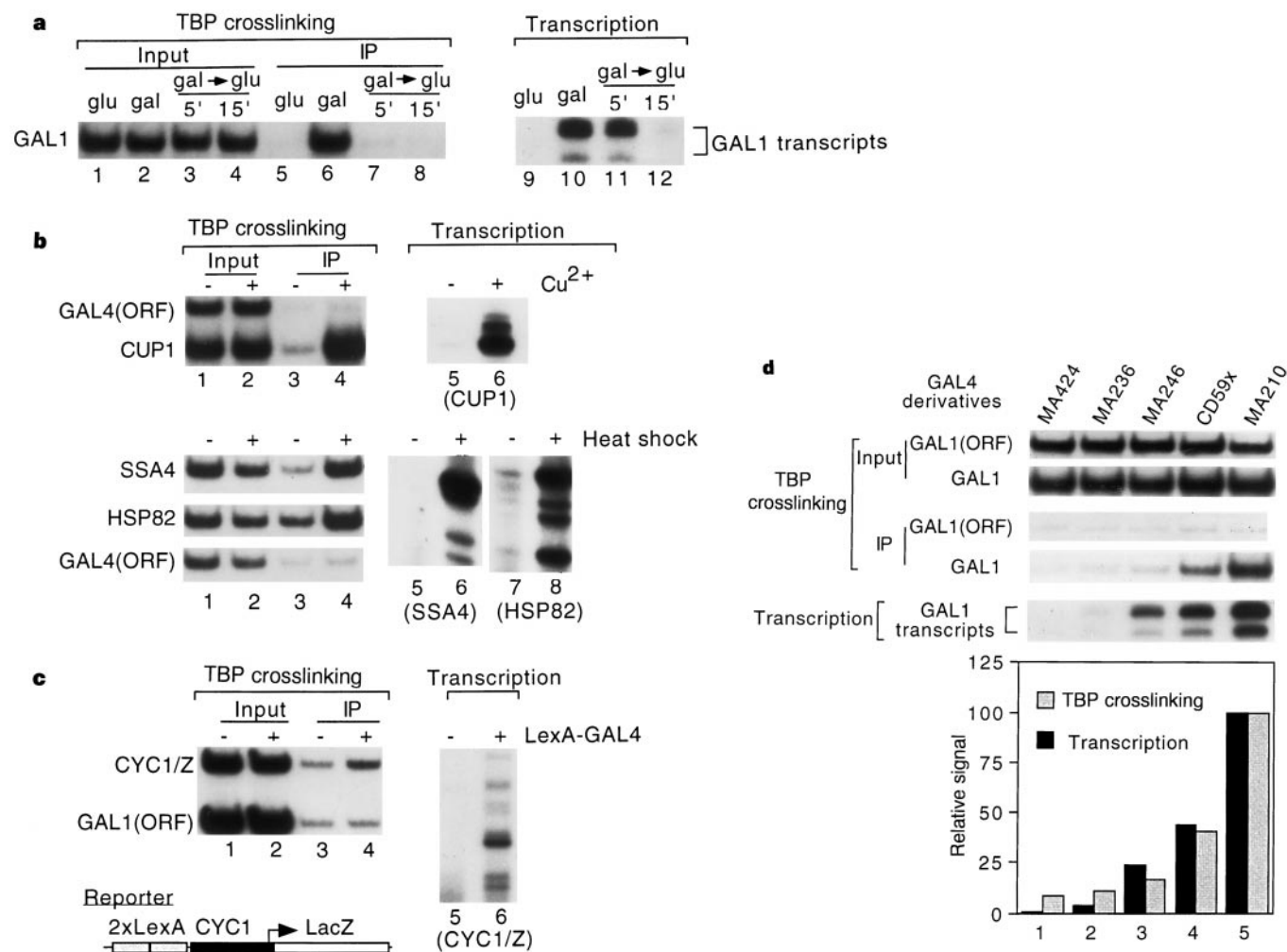


Figure 2 Correlation between TBP binding and transcriptional activity *in vivo*. **a**, *GAL1* transcriptional shut-off. TBP crosslinking was carried out at different times (0, 5, 10 min) after glucose (2%) was added to cells grown in galactose-containing medium. In parallel, cells were collected and total cellular RNA was prepared to analyse *GAL1* transcripts by primer extension. **b**, *CUP1*, *SSA4*, *HSP82*. To analyse the *CUP1* promoter, cells were grown in minimal complete medium and induced for 15 min by addition of CuSO₄ to a final concentration of 1 mM. For *SSA4* and *HSP82*, TBP-crosslinking was analysed immediately before or 15 min after cells were shifted to 39 °C. **c**, *CYC1*. Yeast transformants containing

the reporter, pCTLx⁸, and the LexA-GAL4 expression plasmid or control vector, were grown in minimal selective medium. TBP binding to the hybrid promoter was detected using a primer within the *CYC1* sequence and another at the junction of *CYC1*-*LacZ* fusion. The latter primer was also used in primer-extension analysis of the *CYC1*/Z transcripts. **d**, Activator series. GY1 cells (*gal1*⁻ *gal180*⁻) transformed with plasmids expressing different GAL4-deletion derivatives¹⁰ were grown in minimal selective medium containing 2% glycerol, 2% lactate and 2% galactose. TBP-crosslinking to the *GAL1* promoter and, as a control, to a sequence within the *GAL1* open reading frame (ORF) was analysed.

interaction between the TATA-box-binding protein (TBP) and its binding site, the TATA box. Analysis of a variety of endogenous yeast genes, and of a series of activators of differing strength, reveals a general correlation between TBP binding and transcriptional activity. Using mutant yeast strains, we show that Mot1 prevents the binding of TBP to inactive promoters and that activator-mediated stimulation of TBP binding requires additional GTFs, including TFIIB and Srb4. Taken together, our results indicate that TBP binding *in vivo* is stringently controlled, and that the ability of activators to stimulate this step in the assembly of the preinitiation complex is a highly cooperative process involving multiple transcription factors.

DNA footprinting has been used to study the relation between TBP binding and transcriptional activation *in vivo*⁴⁻⁶, but the conclusions of these studies are inconsistent, probably because of the technical difficulties of these procedures and the uncertainty of whether a TATA-box footprint is actually due to TBP binding. To analyse TBP binding to promoters *in vivo*, we adopted a formaldehyde crosslinking/immunoprecipitation technique^{7,8}. In brief, TBP-DNA crosslinks were induced by addition of formaldehyde to living yeast cells, TBP-DNA complexes were immunopurified using an anti- γ TBP antibody, and specific DNA fragments in the immunoprecipitate were quantified by using the polymerase chain reaction (PCR).

Figure 1a shows the three DNA fragments that were analysed: the constitutively transcribed *ADH1* promoter, the galactose-inducible *GAL1* promoter and, as a control, a region within the *GAL4* protein-coding sequence (*GAL4* ORF), which lacks a promoter. The results indicate that significant and roughly comparable levels of TBP were bound to the *ADH1* promoter when cells were grown in either glucose or galactose, whereas for the *GAL1* promoter, TBP binding was undetectable when cells were grown in glucose but increased

dramatically following growth in galactose (Fig. 1a, lanes 7 and 8). These binding results correlated precisely with the transcriptional activity of these genes.

Several controls in Fig. 1a confirmed that the *in vivo* crosslinking assay provided a valid measure of TBP binding: the signal was dependent both upon addition of the crosslinker formaldehyde (compare lanes 2 and 8) and immunoprecipitation with an anti- γ TBP antibody (compare lanes 7, 8 and 9, 10). In addition, only a weak signal was detected for the *GAL4* DNA fragment (lanes 7 and 8), which lacks a promoter and provided a measure of the assay background in the experiments presented below. Most important, mutation of the *GAL1* TATA box⁴ dramatically reduced TBP crosslinking to background levels (Fig. 1b).

When cells were grown in galactose, *GAL1* was actively transcribed, but when glucose was added, transcription rapidly ceased (Fig. 2a, lanes 10-12). Figure 2a shows that the transcriptional shut-down is accompanied by a loss of TBP binding (lanes 6-8). Figure 2b analyses the copper-inducible *CUP1* gene and two heat-shock genes, *HSP82* and *SSA4*: for *CUP1*, transcription and TBP binding were, once again, well correlated (compare lanes 3, 4 and 5, 6). For *SSA4* and *HSP82*, transcription substantially increased after heat-shock (lanes 5-8), as did TBP binding (lanes 3, 4).

Earlier *in vivo* footprinting studies indicated that TBP was constitutively bound to the *CYC1* promoter, regardless of its transcriptional status⁶. In Fig. 2c, we analyse TBP binding to a *CYC1* promoter derivative containing two upstream LexA-binding sites⁹. The results indicate that upon addition of the activator, TBP binding modestly increased (compare lanes 3 and 4), but was substantially less than the transcriptional increase (compare lanes 5 and 6), which is consistent with the previous study⁹.

To quantify the relation between TBP binding and transcriptional activity, we analysed a series of *GAL4* fusion proteins bearing activation regions of different strengths¹⁰. Figure 2d shows that there is a strong correlation between TBP binding and transcription.

To determine whether activators affected TBP and other GTFs comparably, we used our crosslinking assay to measure the association of TFIIB and RNA polymerase II with promoters. Figure 3a shows that, for all promoters tested (*ADH1*, *GAL1*, *HSP82*, *SSA4* and *CUP1*), there was a similar correlation between transcriptional

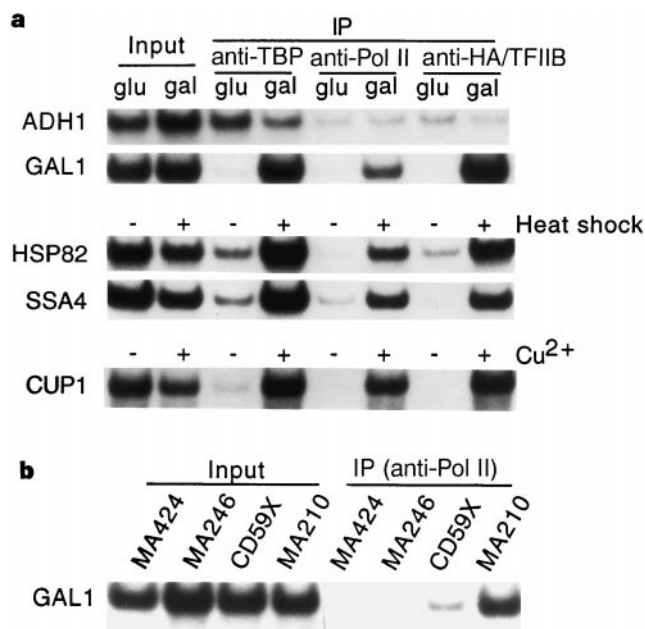


Figure 3 Correlation between transcriptional activity and association of TFIIB and RNA polymerase II with promoters. **a**, *ADH1*, *GAL1*, *CUP1*, *HSP82* and *SSA4*. The yeast strain used in this analysis lacks the endogenous gene encoding TFIIB (*SUA7*) and contains a plasmid expressing TFIIB with an N-terminal triple influenza haemagglutinin (HA) epitope tag. Cell growth and induction were done as described for Figs 1, 2. Immunoprecipitations were done with antibodies against γ TBP, the mouse monoclonal antibody 8WT16 (BAbCO) against the CTD domain of the RNA polymerase II large subunit, or the mouse monoclonal antibody 16B12 (BAbCO) against the HA epitope tag. **b**, Activator series. The same as Fig. 2d, except that the 8WG16 antibody directed against RNA polymerase II was used.

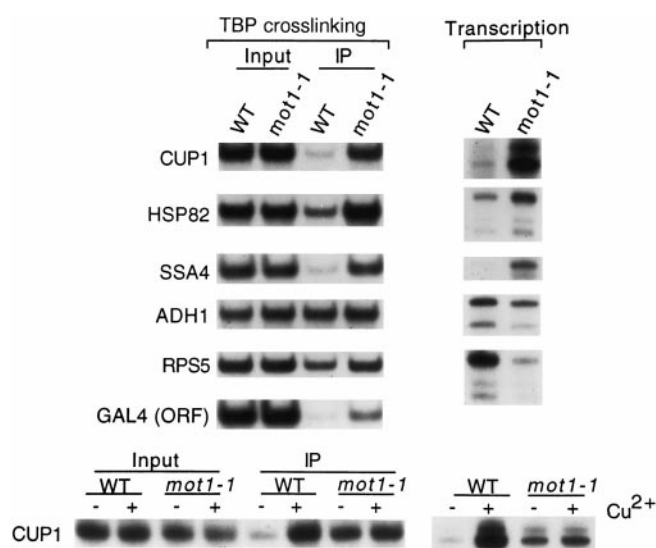


Figure 4 Mot1 prevents TBP binding to inactive promoters. Wild-type (JD195) or mutant *mot1-1* (JD215B) cells¹² were shifted from 30°C to 37°C for 1 h before TBP-crosslinking and RNA analysis. For the experiment shown in the bottom panel, CuSO_4 was added to the culture 1 h after the temperature shift and the cells were incubated for an additional 30 min. Comparable results were obtained at 30°C and 37°C (data not shown).

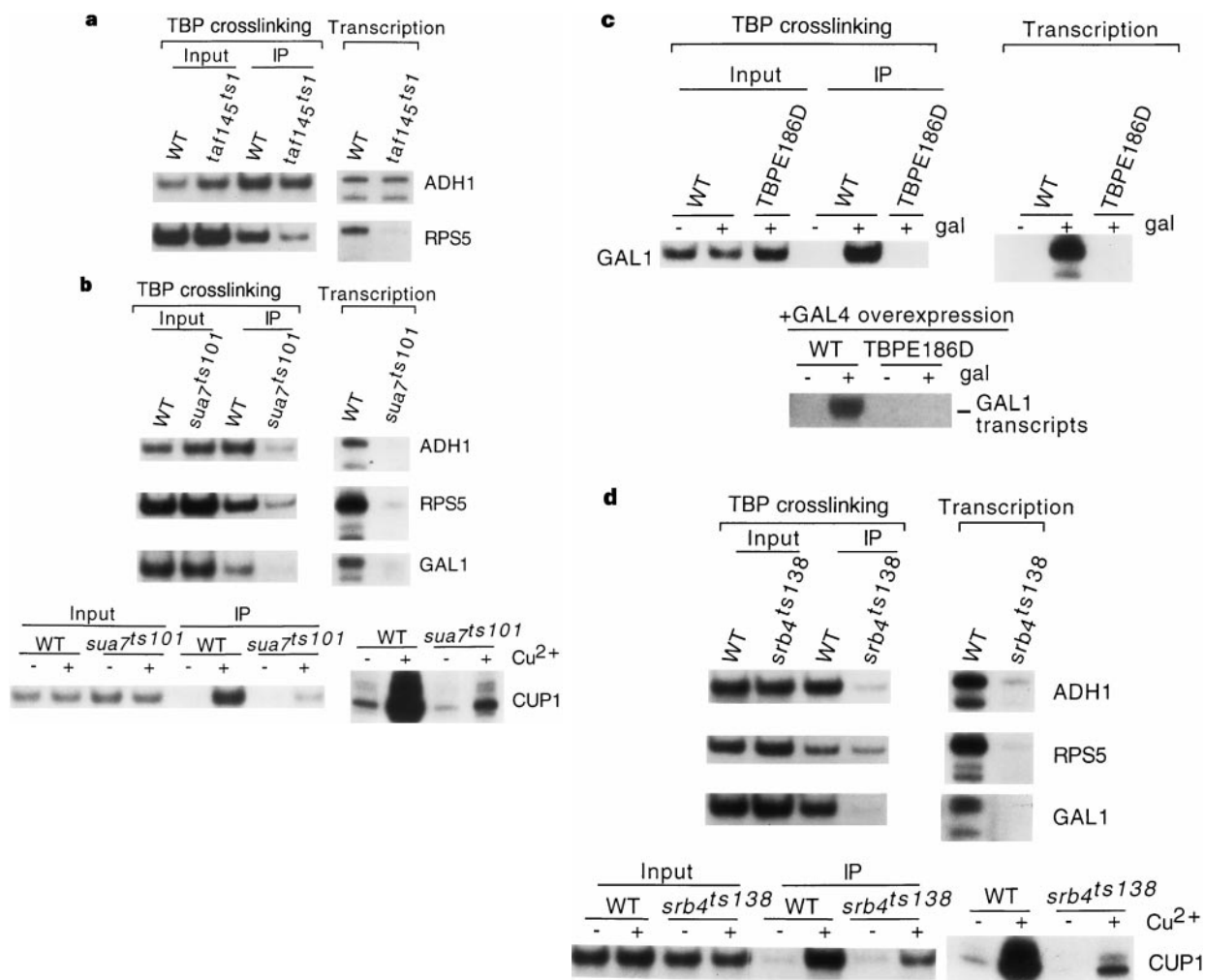


Figure 5 Transcription components that facilitate activator-mediated stimulation of TBP binding. **a**, TAF_{II}145. The TAF_{II}145 temperature-sensitive mutant¹⁵ and the corresponding isogenic wild-type cells were grown at 24°C and shifted to 38°C for 1 h before analysis of transcription or TBP binding. **b**, TFIIB. The same as in **a**, except that for analysis of the *GAL1* promoter, cells were grown in YP galactose. For the *CUP1* promoter, cells were grown in minimal complete medium and

induced by CuSO₄ 30 min after the temperature shift. **c**, TBPE186D. Cells expressing TBPE186D were grown in glucose-containing medium at 30°C. For *GAL1* induction, cells were pelleted and resuspended in galactose-containing medium. To overexpress GAL4, cells were transformed with pMA210, which expresses GAL4 under the control of the *ADH1* promoter¹⁰. **d**, *Srb4*. As for **b**, except that the *Srb4*^{ts138} mutant²² was analysed.

activity and promoter occupancy for TBP, TFIIB and RNA polymerase II. The results shown in Fig. 3b reveal that for RNA polymerase II, as with TBP (Fig. 2d), there was also a strong correlation between the strength of the activation region and promoter occupancy.

Various protein factors have been implicated as general transcription repressors, and one of these, Mot1, can modulate TBP binding *in vitro*¹¹, although it has not been determined whether Mot1 modulates TBP binding *in vivo*. Figure 4 shows that upon inactivation of Mot1 (ref. 12), TBP binding to all promoters (*CUP1*, *HSP82*, *SSA4*, *ADH1* and *RPS5*), as well as to the *GAL4* protein-coding sequence, increased. Thus, Mot1 has a general ability to prevent TBP binding *in vivo*. Although TBP binding increased in all cases, inactivation of Mot1 increased transcription of some genes (*HSP82*, *SSA4*, and *CUP1* in the absence of Cu²⁺), but had little effect on *ADH1* transcription and, unexpectedly, transcription from *CUP1*, in the presence of Cu²⁺, and *RPS5* decreased. These diverse transcriptional effects following Mot1 inactivation are consistent with previous results^{13,14}.

We next examined the role of other transcription components in the activator-mediated stimulation of TBP binding. We analysed three factors: TAF_{II}145, a highly gene-specific TFIID component

that functions as a core promoter-selectivity factor¹⁵; TFIIB, a GTF that enters the preinitiation complex (PIC) immediately after TBP^{1,2}; and *Srb4*, a GTF that is tightly associated with the carboxy-terminal domain (CTD) of RNA polymerase II (refs 16, 17). Figure 5a shows, as expected¹⁵, that TAF_{II}145 is required for transcription of *RPS5* but not of *ADH1*. Consistent with these transcription results, following temperature-sensitive inactivation of TAF_{II}145, binding of TBP to *RPS5* decreased significantly, whereas *ADH1* was relatively unaffected. These results suggest that TAF_{II}145 facilitates TBP binding in a promoter-specific fashion.

Consistent with its role as a GTF, temperature-sensitive inactivation of TFIIB led to a decrease in transcription of all genes tested. In all cases, TBP binding also decreased substantially (Fig. 5b). In a complementary experiment, we analysed a TBP missense mutant (E186D) that contains a substitution in one of the three residues contacted by TFIIB^{18–21}. As a result of this substitution, the TBP–TFIIB interaction is severely weakened and yeast cells expressing the TBPE186D mutant exhibit a slow-growth phenotype and multiple gene-specific activation defects (Fig. 5c, and our unpublished data). Figure 5c shows that yeast expressing TBPE186D cannot support galactose-inducible *GAL1* transcription either when transcription is directed by endogenous *GAL4* or, to ensure adequate amounts of

GAL4, when GAL4 is overexpressed from the *ADHI* promoter. GAL4 also failed to stimulate binding of TBPE186D to the *GALI* promoter, explaining the transcription defect and directly demonstrating the importance of the TBP–TFIIB interaction for activator-mediated stimulation of TBP binding.

Srb4 is required for transcription in general²², and has been implicated as a direct target of certain activators²³. As expected for a GTF, Srb4 was required for *ADHI*, *RPS5*, *GALI* and *CUP1* transcription (Fig. 5d). Following inactivation of Srb4, there was also a loss of TBP binding to the *GALI*, *ADHI*, *CUP1* and, to a lesser extent, the *RPS5* promoters.

On the basis of these results, we infer the following. First, there is a general qualitative and quantitative correlation between transcriptional activity and the association of TBP, as well as other GTFs, with the promoter. Thus, *in vivo* activators function, at least in part, by recruiting GTFs and ultimately RNA polymerase II, as proposed on the basis of *in vitro* assays measuring promoter–GTF interactions²⁴, and more recently inferred by ‘activator bypass’ experiments³. Second, TBP binding *in vivo* is actively prevented by Mot1. Finally, activator-mediated stimulation of TBP binding requires GTFs other than TFIID and its components. The requirement for multiple GTFs is not readily compatible with models proposing that activators stimulate TBP binding and transcription solely through contact with TFIID^{25,26}. The involvement of multiple GTFs in activator-mediated stimulation of TBP binding underscores the cooperative nature of PIC assembly and is consistent with several possible mechanisms, including a pre-assembled ‘holoenzyme’^{27,28} that is disrupted by inactivation of certain GTFs, or direct contact between the activator and GTFs such as TFIIB²⁴ or Srb4 (ref. 23), which promote TBP binding and PIC assembly through cooperative interactions. □

Methods

Unless otherwise specified, the yeast strain was W303a and the cells were grown in YP or minimal medium containing glucose. The formaldehyde crosslinking/immunoprecipitation method has been described^{7,8}. Briefly, yeast cells were treated with 1% formaldehyde, collected and resuspended in lysis buffer⁷. Following sonication, which generated DNA fragments averaging ~500 bp, cell lysates were precleared by centrifugation; Sarkosyl (1%) was added to the lysate before immunoprecipitation with an anti-yTBP polyclonal antibody⁷. Immunoprecipitated protein–DNA complexes were treated with protease K, the crosslinks reversed, and the DNA purified and analysed by quantitative PCR⁸. PCR reactions contained [α -³²P]dATP (2.5 μ Ci for each 25- μ l reaction) and the PCR products were detected by autoradiography after separation on a 6% polyacrylamide gel. Primers corresponded to sequences around the TATA element, and the PCR products ranged from 130 to 200 bp; the control fragments *GAL4* (ORF) and *GALI* (ORF) are located, respectively, 2,000 and 900 bp downstream of the TATA sequences of their promoters. To analyse transcription, total cellular RNA was prepared and primer extension was carried out using primers close to the RNA cap site.

Received 3 December 1998; accepted 8 April 1999.

- Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
- Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
- Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
- Selleck, S. B. & Majors, J. *In vivo* ‘photofootprint’ change at sequences between the yeast *GALI* upstream activating sequence and ‘TATA’ element required activated *GALI* protein but not a functional TATA element. *Proc. Natl Acad. Sci. USA* **85**, 5399–5403 (1988).
- Giardina, C. & Lis, J. T. Dynamic protein–DNA architecture of a yeast heat shock promoter. *Mol. Cell Biol.* **15**, 2737–2744 (1995).
- Chen, J., Ding, M. & Pederson, D. S. Binding of TFIID to the *CYC1* TATA boxes in yeast occurs independently of upstream activating sequences. *Proc. Natl Acad. Sci. USA* **91**, 11909–11913 (1994).
- Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11**, 83–93 (1997).
- Orlando, V. & Paro, R. Mapping polcomb-repressed domains in the bithorax complex using *in vivo* formaldehyde cross-linked chromatin. *Cell* **75**, 1187–1198 (1993).
- Lue, N. F., Buchman, A. R. & Kornberg, R. D. Activation of yeast RNA polymerase II transcription by a thymidine-rich upstream element *in vitro*. *Proc. Natl Acad. Sci. USA* **86**, 486–490 (1989).
- Ma, J. & Ptashne, M. Deletion analysis of *GAL4* defines two transcriptional activating segments. *Cell* **48**, 847–853 (1987).

- Auble, D. T. *et al.* Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.* **8**, 1920–1934 (1994).
- Davis, J. L., Kunisawa, R. & Thorne, J. A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in the yeast *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **12**, 1879–1892 (1992).
- Collart, M. A. The NOT, SPT3, and MOT1 genes functionally interact to regulate transcription at core promoters. *Mol. Cell Biol.* **16**, 6668–6676 (1996).
- Madison, J. M. & Winston, F. Evidence that Spt3 functionally interacts with Mot1, TFIIA, and TATA-binding protein to confer promoter-specific transcription control in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **17**, 287–295 (1997).
- Shen, W.-C. & Green, M. R. Yeast TAF_{II}145 functions as a core promoter selectivity factor, not a general coactivator. *Cell* **90**, 615–624 (1997).
- Thompson, C. M., Koleske, A. J., Chao, D. M. & Young, R. A. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361–1375 (1993).
- Kim, Y.-J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
- Nikolov, D. B. *et al.* Crystal structure of a TFIIB–TBP–TATA-element ternary complex. *Nature* **377**, 119–128 (1995).
- Tang, H., Sun, X., Reinberg, D. & Ebright, R. H. Protein–protein interactions in eukaryotic transcription initiation: structure of the preinitiation complex. *Proc. Natl Acad. Sci. USA* **93**, 1119–1124 (1996).
- Bryant, G. O., Martel, L. S., Burley, S. K. & Berk, A. J. Radical mutations reveal TATA-box binding protein surfaces required for activated transcription *in vitro*. *Genes Dev.* **10**, 2491–2504 (1996).
- Lee, M. & Struhl, K. A severely defective TATA-binding protein–TFIIB interaction does not preclude transcription activation *in vivo*. *Mol. Cell Biol.* **17**, 1336–1345 (1997).
- Thompson, C. M. & Young, R. A. General requirement for RNA polymerase II holoenzymes *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 4587–4590 (1995).
- Koh, S. S., Ansar, A. Z., Ptashne, M. & Young, R. A. An activator target in the RNA polymerase II holoenzyme. *Mol. Cell* **1**, 859–904 (1998).
- Lin, Y. S. & Green, M. R. Mechanism of action of an acidic transcriptional activator *in vitro*. *Cell* **64**, 971–981 (1991).
- Burley, S. K. & Roeder, R. G. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**, 768–799 (1996).
- Verrijzer, C. P. & Tjian, R. TAFs mediate transcriptional activation and promoter selectivity. *Trends Biochem. Sci.* **21**, 338–342 (1996).
- Koleske, A. J. & Young, R. A. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* **20**, 113–115 (1995).
- Halle, J.-P. & Meisterernst, M. Gene expression: increasing evidence for a transcriptosome. *Trends Genet.* **12**, 161–163 (1996).

Acknowledgements. We thank R. Young, M. Grunstein, J. Thorne and J. Majors for yeast strains, J. Ma for plasmids, and K. Struhl for discussion and for communicating unpublished results. This work was supported in part by a grant from the NIH to M.R.G. M.R.G. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to M.R.G. (e-mail: michael.green@ummedu.edu).

Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme

Laurent Kuras & Kevin Struhl

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

In eukaryotes, transcriptional activators have been proposed to function by recruiting the RNA polymerase II (Pol II) machinery^{1–3}, by altering the conformation of this machinery^{4,5}, or by affecting steps after initiation^{6–8}, but the evidence is not definitive. Genomic footprinting of yeast TATA-box elements reveals activator-dependent alterations of chromatin structure⁹ and activator-independent protection¹⁰, but little is known about the association of specific components of the Pol II machinery with promoters *in vivo*. Here we analyse TATA-box-binding-protein (TBP) occupancy of 30 yeast promoters *in vivo*. We find that TBP association with promoters is stimulated by activators and inhibited by the Cyc8–Tup1 repressor, and that transcriptional activity correlates strongly with the degree of TBP occupancy. In a small subset of promoters, TBP occupancy is higher than expected when gene activity is low, and the activator-dependent increase is modest. TBP association depends on the Pol II holoenzyme component Srb4, but not on the Kin28 subunit of the transcription factor TFIIF, even though both proteins are generally required for transcription. Thus in yeast cells, TBP association with promoters occurs in concert with the Pol II holoenzyme, activator-dependent

recruitment of the Pol II machinery occurs at the vast majority of promoters, and Kin28 acts after the initial recruitment.

We analysed TBP occupancy of promoters in yeast cells by using a modified version of a chromatin immunoprecipitation procedure^{11,12}. Initially, we analysed the *PGK1* promoter and genomic regions that are centrally located within large (5 kilobase (kb)) structural genes (*RPO21* and *POL1*) and hence as far as possible from promoters. As expected, the product of the polymerase chain reaction (PCR) corresponding to the *PGK1* promoter is easily detected in cells containing the epitope-tagged TBP but not in control cells containing non-tagged TBP (Fig. 1). Crosslinking efficiency at the *PGK1* promoter is 0.9% (hereby defined as 10 units; see Methods), which is in accord with results obtained with other yeast DNA-binding proteins^{12,13}. TBP appears to cross-link at low efficiency to the *RPO1*, *POL1*, and *GLT1* (results not shown) structural genes (compare strains containing tagged rather than non-tagged TBP). This observation suggests that TBP weakly and non-specifically associates with most (perhaps all) chromosomal DNA *in vivo*.

To investigate transcriptional activation mechanisms, we determined TBP occupancy of various promoters that respond to specific DNA-binding activators. Gal4-dependent promoters are transcriptionally active in cells grown in galactose, but not in glucose or raffinose. TBP strongly associates with all four *GAL* promoters tested (40–60 units) in galactose-grown cells, but is not detectably associated in glucose- or raffinose-grown cells (Fig. 1a). Strong TBP occupancy requires a functional promoter, as evidenced by the weak signal corresponding to *UAS_G*, the region just upstream of the *GAL7* enhancer. These results extend the previous observation that photofootprinting *in vivo* detects Gal4- and transcription-dependent changes at the *GAL1* and *GAL10* TATA elements⁹. Similarly, increased TBP occupancy is observed at four *MET* promoters that are activated by heteromeric complexes containing the Met4 activator when cells are grown in medium lacking methionine¹⁴ (Fig. 1b). Thus, Gal4 and Met4 stimulate TBP recruitment at several different core promoters.

Yeast cells exposed to a transient heat shock rapidly induce transcription of various genes using distinct molecular mechanisms. For genes directly activated by Hsf1 (*SSA3*, *SSA4*, *HSP82*, and probably *HSP104*) or the redundant Msn2 and Msn4 activators (*CTT1*, *HSP12*, and probably *HSP104*), heat shock dramatically increases TBP occupancy *in vivo* (Fig. 2a). In addition, heat shock causes increased transcription (Fig. 2c) and increased TBP occupancy of the glycolytic genes *ENO1* and *PGK1*, but not of *PYK1*. Finally, in another rapid stress response that involves the Ace1 activator, copper-dependent induction of *CUP1* transcription (40-fold; data not shown) is associated with a concomitant 40-fold increase in TBP occupancy at the *CUP1* promoter (Fig. 2b). Taken together, these results demonstrate that Ace1, Hsf1, Msn2, Msn4, and unknown activators increase TBP recruitment, and they strengthen the correlation between transcriptional activity and TBP occupancy.

The transition between growth on glucose and non-fermentable carbon sources such as ethanol requires the induction of many genes, including those responsive to Adr1, Hap1 and unknown activators. In accord with their transcriptional activities, TBP is strongly associated with the *ADH2*, *ICL1*, *PCK1*, and *CYB2* promoters when cells are grown in ethanol medium, but not when cells are grown in glucose medium (Fig. 3). In addition, *PGK* transcription is reduced threefold in ethanol medium, and TBP occupancy decreases by a comparable extent.

In contrast to the other 28 promoters tested, TBP clearly associates with the *CYC1* and *COX5a* promoters in glucose medium even though transcription levels are very low (Fig. 3). Moreover, TBP occupancy at these promoters is barely stimulated (less than twofold) when cells are grown in ethanol medium, even though transcription is stimulated by a factor of 10. These results are partially consistent with a DNase I footprint over the *CYC1* TATA element in isolated nuclei, even under conditions where transcription was not observed¹⁰. However, DNase I footprints imply a high degree of protein occupancy (at least 50%), whereas TBP occupancy at the *CYC1* promoter is considerably lower (5–10%; see below);

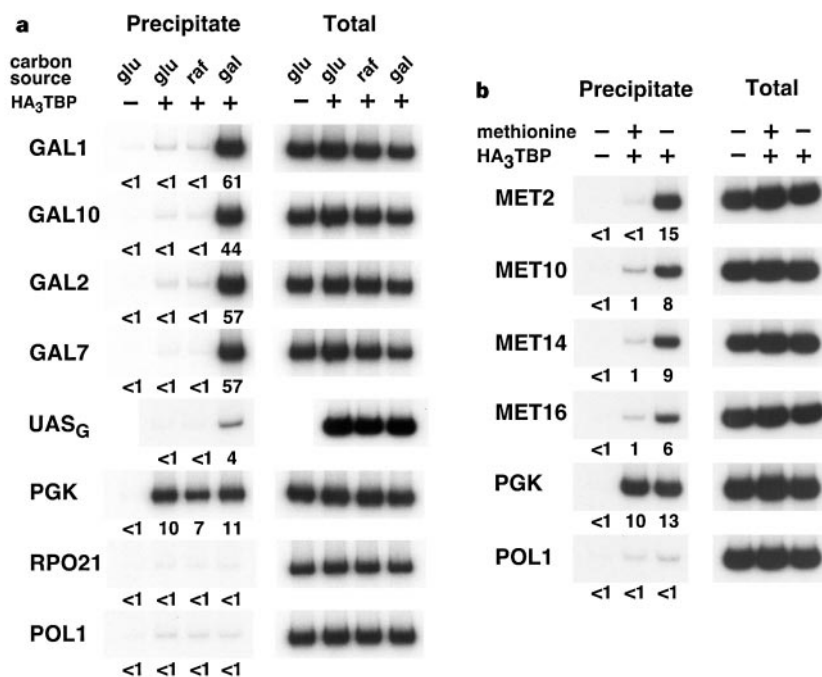


Figure 1 TBP occupancy at *GAL* and *MET* promoters. **a**, Cells containing HA₃-tagged or untagged TBP were grown in: **a**, YP medium containing glucose (glu), raffinose (raf) or galactose (gal) as carbon source; or **b**, glucose minimal medium in the presence or absence of 1 mM methionine. PCR products corresponding to the indicated promoters, the *RPO21* and *POL1* structural genes, or the region

just upstream from *UAS_G* at *GAL7* were generated from total chromatin or immunoprecipitated DNA. The amount of immunoprecipitated DNA assayed by PCR varied slightly depending on the promoter examined, and this was considered in determining units of TBP occupancy (indicated under the corresponding panels).

hence, it is unclear whether the previously observed DNase I footprint corresponds to TBP association.

Finally, we examined whether the Cyc8–Tup1 co-repressor exerts its effects at the level of TBP occupancy. The Cyc8–Tup1 complex is recruited by pathway-specific DNA-binding proteins to promoters that are regulated by cell-type (*MFA1*), glucose (*SUC2*) and oxygen (*ANB1*)^{15,16}. In wild-type strains, *MFA1*, *ANB1* and *SUC2* are essentially not expressed, and TBP occupancy occurs at background level (Fig. 4). However, in a *tup1* strain, transcription of these genes occurs to various extents, and TBP associates with the promoters. Thus, Cyc8–Tup1 inhibits TBP association with three promoters that respond to different DNA-binding repressors and activators and contain unrelated core promoter regions.

The maximal level of TBP occupancy (50–100 units) is observed for several highly active genes, and hence is likely to correspond to full occupancy of the promoter by TBP. This suggests that our measurements reflect TBP occupancy *per se* and so are not generally affected by conformational differences at TATA elements that might affect crosslinking efficiency. In support of this, TBP occupancy at all five transfer-RNA promoters tested occurs at this maximal value (our unpublished results), even though TBP plays distinct roles at Pol II and Pol III promoters. Moreover, the TATA-element sequence has no significant effect on the crystal structure of the TBP–TATA complex¹⁷. These considerations, and the very strong correlation between transcription levels and TBP occupancy, strongly suggest that we have quantitatively measured TBP occupancy (estimated experimental error is $\pm 30\%$).

The promoters analysed here differ in the quality, location and number of activator-binding sites and TATA elements, as well as in nucleosome position and phasing. Despite these differences, our results show that the predominant, although not exclusive, limitation for transcriptional activity in wild-type cells correlates with the ability of TBP to associate stably with the promoter. As the remainder of the Pol II machinery cannot associate with promoters

in the absence of TBP (or TFIID), our results strongly suggest that the entire Pol II machinery is absent from the vast majority of promoters in the absence of functional activator proteins, although they do not indicate which components of the Pol II machinery are important for TBP recruitment or whether partial preinitiation complexes can stably associate with promoters *in vivo*.

To investigate these issues, we examined TBP occupancy under conditions in which Pol II transcription was generally reduced by temperature-sensitive mutations in the following components: *Srb4*, a component of the mediator and Pol II holoenzyme complexes¹⁸; *Rpb1*, the largest subunit of Pol II (ref. 19); and *Kin28*, the subunit of TFIIF that phosphorylates the C-terminal tail of Pol II (ref. 20) (Fig. 5). Unexpectedly, loss of *Srb4* function significantly reduces TBP occupancy at all six promoters tested. TBP occupancy is also decreased in the *rpb1* mutant strain, although the effect is less marked and not seen at all promoters. In contrast, TBP occupancy is not affected upon inactivation of *Kin28*, even though the reduction in transcription is comparable to that from the *rpb1* and *srb4* strains. Thus, TBP association with promoters requires the Pol II holoenzyme, but not the *Kin28* subunit of TFIIF.

The *Srb4* requirement for TBP occupancy strongly suggests that

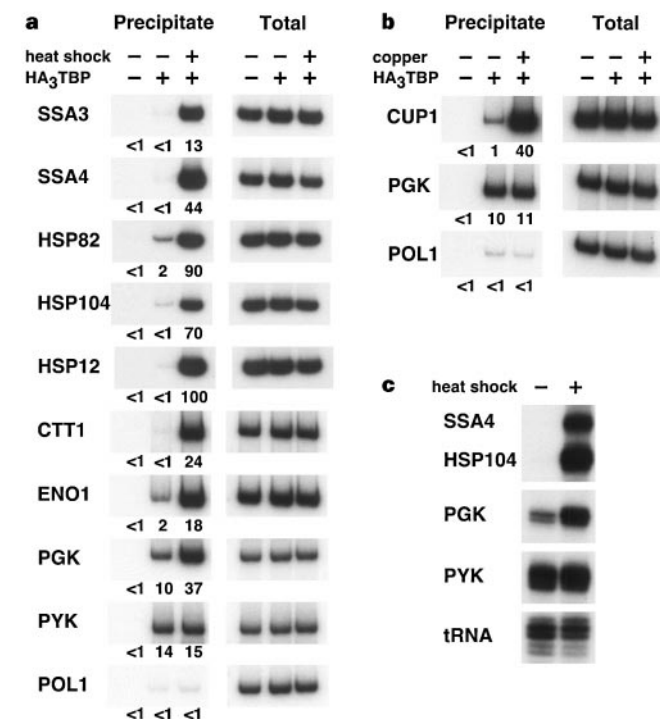


Figure 2 TBP occupancy in response to heat shock and copper induction. Cells containing HA₃-tagged or untagged TBP were grown in glucose minimal medium at 24 °C and subjected to: **a**, a 15-min heat shock at 39 °C; or **b**, a 20-min treatment with 1 mM copper sulphate. TBP occupancy was analysed as for Fig. 1. **c**, Quantitative S1 analysis.

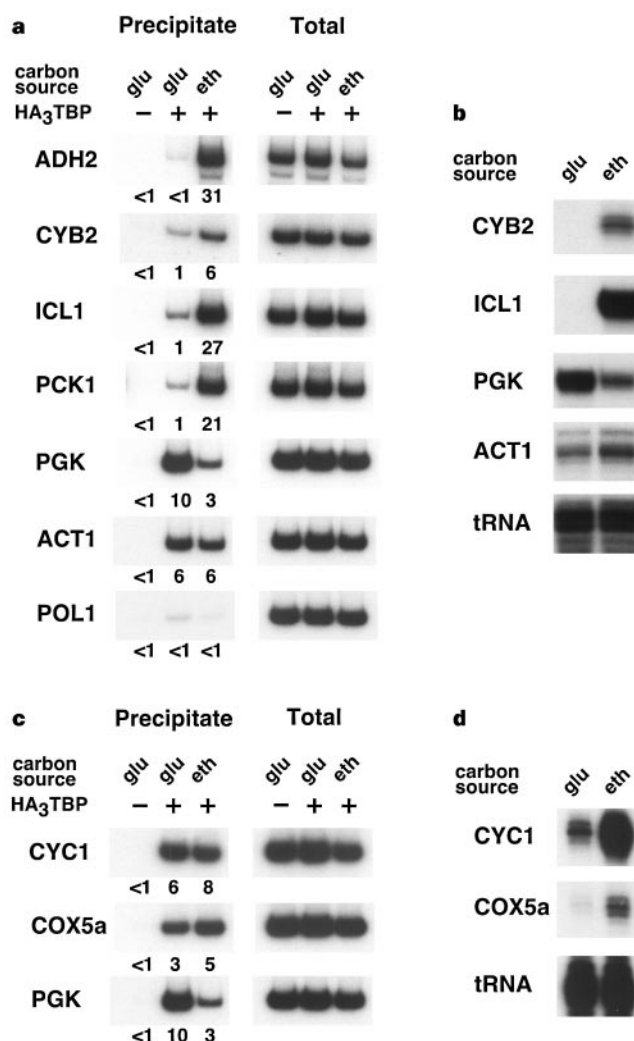


Figure 3 TBP occupancy at promoters induced during non-fermentative growth. **a**, **c**, Cells containing HA₃-tagged or untagged TBP were grown in synthetic complete medium containing 4% glucose as carbon source. Half of the culture was washed with medium lacking glucose and then transferred to medium containing 3% ethanol for 6 h (eth). TBP occupancy was analysed as for Fig. 1. **b**, **d**, Quantitative S1 analysis.

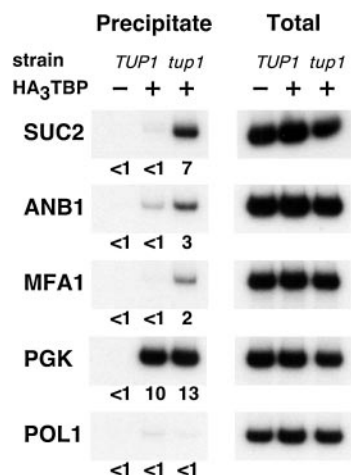


Figure 4 TBP occupancy at promoters repressed by the Cyc8-Tup1 corepressor complex. *TUP1* or *tup1Δ::LEU2* cells containing HA₃-tagged or untagged TBP were grown in YPD medium, and TBP occupancy was analysed as for Fig. 1.

the association of TFIID (the biologically relevant form of TBP) and the Pol II holoenzyme are concerted events. Although these events might be mechanistically and temporally distinct, it appears that neither TFIID nor the Pol II holoenzyme can individually associate stably with promoters *in vivo*. As a corollary, our results argue that biochemically stable subcomplexes (such as TBP-TFIIB-TATA), as well as functional preinitiation complexes corresponding to the minimal transcription machinery, are relatively unstable *in vivo*. As such, our results favour (although they do not prove) the idea that preinitiation complexes are formed by recruitment of a pre-assembled Pol II holoenzyme, rather than by stepwise recruitment of individual factors.

The dispensability of the Kin28 subunit of TFIIF for TBP occupancy suggests that, *in vivo*, phosphorylation of the C-terminal tail of Pol II occurs after the concerted (and presumably stable) association of TFIID and the Pol II holoenzyme. This would be consistent with the observations that Kin28-dependent phosphorylation of the Pol II tail occurs only on a preassembled complex²¹ and that formation of the preinitiation complex requires unphosphorylated Pol II (ref. 22). Our results do not address whether TFIIF association occurs together with or subsequent to the association of the Pol II holoenzyme, although we note that TFIIF is not a component of the Pol II holoenzyme^{23,24} and is not absolutely required for transcription under certain biochemical conditions²². The modest decrease in TBP occupancy in the *rpb1* strain is hard to interpret, although it suggests that the mutated Pol II is partially functional in preinitiation complex formation and defective, to some extent, at a later step such as promoter clearance or elongation.

Our results constitute direct physical evidence that the predominant mechanism by which activators stimulate (and repressors inhibit) transcription in yeast cells involves recruitment of the Pol II machinery to promoters. At more than 90% of the natural promoters we examined, there appear to be two states: a transcriptionally inactive state, in which TBP and hence the entire Pol II machinery are not associated, and an activator-dependent state, in which the level of transcription correlates with the amount of TBP occupancy. However, the concerted association of TFIID and the Pol II holoenzyme with promoters makes it impossible to determine whether activators directly recruit TFIID, Pol II holoenzyme, or both. In addition, activators could function by recruiting chromatin-modifying activities, thereby generating an altered chromatin structure that increases the association of TBP and Pol II holoenzyme. Finally, although recruitment is the predominant mechanism, TBP

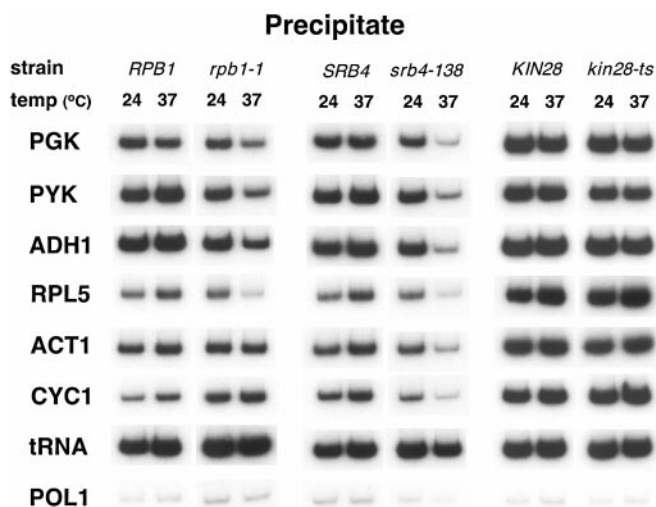


Figure 5 Role of general transcription factors for TBP occupancy at promoters. Isogenic wild-type or the indicated mutant cells (all containing untagged TBP) were grown in YPD medium and shifted to 37 °C, the restrictive temperature, for 1 h (or 75 min in the case of the *kin28* strain). PCR products corresponding to the indicated promoters, including tRNA, which responds to Pol III (or *POL1* structural gene), were generated from total chromatin (not shown) or DNA immunoprecipitated with antibodies against TBP.

occupancy is not strictly limiting for activation of *CYC1* and *COX5a*. It may be that, in a small subset of promoters, the transcriptionally inactive state loosely resembles the situation in Kin28-deficient cells, in which it is presumed that TFIID and Pol II holoenzyme are associated, but a late step such as TFIIF recruitment or promoter clearance is defective; in this regard, TFIIF is associated with the *CYC1* and *COX5a* promoters in glucose medium (P. Kosa and K.S., unpublished observations). Although the detailed mechanisms remain to be determined, our findings explain observations suggesting that activators can function at distinct steps in yeast cells²⁵⁻²⁷. □

Methods

Yeast strains. The haemagglutinin (HA)-tagged TBP molecule contains three copies of the HA epitope inserted after codon 3 of the TBP open reading frame. For the experiment in Fig. 1, *URA3* centromeric plasmids expressing TBP or HA₃-TBP from the TBP promoter were introduced into BYΔ2 by plasmid shuffling²⁸. Otherwise, *URA3* integrating plasmids containing N-terminal portions (residues 1–172) of TBP or HA₃-TBP were introduced into the *TBP* locus of W303-1A (Figs 2, 3) or FT5 (ref. 16) (Fig. 4) by integrative disruption. For the experiment in Fig. 5, the isogenic *RBP1* and *rpb1-1* (ref. 19) and *KIN28* and *kin28-ts16* (ref. 20) strains have been described; the *SRB4* and *srb4-138* alleles¹⁸ were present on centromeric plasmids in a strain deleted for the chromosomal locus.

Chromatin immunoprecipitation. Chromatin was prepared as described¹¹, with several modifications. Cells (400 ml) grown to an absorbance at 600 nm (*A*₆₀₀) of ~0.5 were treated with 1% formaldehyde for 20 min at room temperature, with occasional swirling. Glycine was added to a final concentration of 330 mM and the incubation continued for 5 min. Cells were collected, washed twice with cold TBS (20 mM Tris-HCl, pH 7.5, 150 mM NaCl), once with cold FA-lysis buffer (50 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% sodium deoxycholate, 0.1% SDS, 1 mM PMSF), and resuspended in 1 ml cold FA-lysis buffer containing 0.5% SDS. An equal volume of glass beads (of diameter 0.5 mm) was added, and cells were disrupted by vortexing for 15 min on ice. The lysate was diluted into 8 ml FA-lysis buffer, and the glass beads were discarded as described¹¹. The crosslinked chromatin was pelleted by centrifugation at 200,000g for 20 min, washed for 1 h with FA-lysis buffer, resuspended in 1.5 ml FA-buffer for 1 h at 4 °C, and sonicated to yield an

average DNA fragment size of 350 base pairs (bp) (range, 100–850 bp). Finally, the sample was adjusted to 8 ml with FA-lysis buffer, clarified by centrifugation at 200,00g for 20 min, and aliquots of the resultant chromatin solution were stored at -80°C .

Chromatin solution (800 μl), adjusted to 275 mM NaCl, was incubated with 10 μl anti-HA monoclonal antibody (Figs 1–4) or 25 μl anti-TBP antibody (Fig. 5) coupled to protein-A–Sephadex beads. After 90 min at room temperature on a rotator, beads were washed twice for 4 min in 1.4 ml FA-buffer, twice in 1.4 ml FA-buffer with 500 mM NaCl, once in 1.4 ml of 10 mM Tris-HCl, pH 8.0, 0.25 M LiCl, 1 mM EDTA, 0.5% N-P40, 0.5% sodium deoxycholate, and once in 1.4 ml TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Immunoprecipitated material was eluted from the beads by heating for 10 min at 65°C in 400 μl of 25 mM Tris-HCl, pH 7.5, 10 mM EDTA, 0.5% SDS. To reverse crosslinks, samples were adjusted to 0.8 mg ml $^{-1}$ Pronase and incubated for 1 h at 42°C and for 5 h at 65°C . After extraction with phenol–chloroform–isoamyl alcohol and chloroform, DNA was ethanol-precipitated overnight at -20°C in the presence of 20 μg glycogen, and resuspended in TE buffer.

Quantitative analysis. Precipitated DNA was analysed by quantitative PCR using primer pairs (24–26mers with $\sim 45\%$ G+C content) for specific regions. PCR was first performed with decreasing amounts of template to determine the linear range for each combination of primer sets and DNA. Typically, 1/100 to 1/500th of the immunoprecipitated DNA and 1/30,000 of the total DNA input were used. Reactions were carried out in 10 μl and contained 1 μM primers, 0.1 mM dNTPs and 0.1 mCi ml $^{-1}$ of ^{32}P -dATP (specific activity, 3,000 Ci mmol $^{-1}$). Cycling was for 90 s at 94°C , followed by 26 cycles (24 cycles in the case of the *CUP1* primers) with 30 s at 94°C , 30 s at 55°C , 1 min at 72°C , then 4 min at 72°C . PCR products (typically, 200–300 bp) were quantified by a Fujix BSA 2040 PhosphorImager. The fraction of immunoprecipitated material for a specific fragment was calculated by dividing the amount of PCR product obtained with immunoprecipitated DNA by the amount obtained with total DNA. The value obtained for the *PGK* promoter (0.9%) was arbitrarily defined as 10 units of promoter occupancy, and this value was used for direct comparison among experiments. We estimate that the overall error is $\pm 30\%$, apart from values below 2 units, where the error is greater.

Received 18 February; accepted 8 April 1999.

1. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–577 (1997).
2. Klein, C. & Struhl, K. Increased recruitment of TATA-binding protein to the promoter by transcriptional activation domains *in vivo*. *Science* **266**, 280–282 (1994).
3. Keaveney, M. & Struhl, K. Activator-mediated recruitment of the RNA polymerase II machinery is the predominant mechanism for transcriptional activation in yeast. *Mol. Cell* **1**, 917–924 (1998).
4. Roberts, S. G. & Green, M. R. Activator-induced conformational change in general transcription factor TFIIB. *Nature* **371**, 717–720 (1994).
5. Chi, T. & Carey, M. Assembly of the isomerized TFIIB-TFIID-TATA ternary complex is necessary and sufficient for gene activation. *Genes Dev.* **10**, 2540–2550 (1996).

6. Rougvie, A. E. & Lis, J. T. Postinitiation transcriptional control in *Drosophila melanogaster*. *Mol. Cell. Biol.* **10**, 6041–6045 (1990).
7. Yankulov, K., Blau, J., Purton, T., Roberts, S. & Bentley, D. L. Transcriptional elongation by RNA polymerase II is stimulated by transactivators. *Cell* **77**, 749–759 (1994).
8. Krumm, A., Hickey, L. B. & Groudine, M. Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation. *Genes Dev.* **9**, 559–572 (1995).
9. Selleck, S. B. & Majors, J. Photofootprinting *in vivo* detects transcription-dependent changes in yeast TATA boxes. *Nature* **325**, 173–177 (1987).
10. Chen, J., Ding, M. & Pederson, D. S. Binding of TFIID to the *CYC1* TATA boxes in yeast occurs independently of upstream activating sequences. *Proc. Natl Acad. Sci. USA* **91**, 11909–11913 (1994).
11. Strahl-Bolsinger, S., Heck, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11**, 83–93 (1997).
12. Aparicio, O. M., Weinstein, D. M. & Bell, S. P. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. *Cell* **91**, 59–69 (1997).
13. Meluh, P. B. & Koshland, D. Budding yeast centromere composition and assembly as revealed by *in vivo* cross-linking. *Genes Dev.* **11**, 3401–3412 (1997).
14. Blaiseau, P. L. & Thomas, D. Multiple transcriptional activation complexes tether the yeast activator Met4 to DNA. *EMBO J.* **17**, 6327–6336 (1998).
15. Keleher, C. A., Redd, M. J., Schultz, J., Carlson, M. & Johnson, A. D. Snf6-Tup1 is a general repressor of transcription in yeast. *Cell* **68**, 709–719 (1992).
16. Tzamarias, D. & Struhl, K. Functional dissection of the yeast Cyc8–Tup1 transcriptional co-repressor complex. *Nature* **369**, 758–761 (1994).
17. Patikoglou, G. & Burley, S. K. Eukaryotic transcription factor–DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 289–325 (1997).
18. Thompson, C. M. & Young, R. A. General requirement for RNA polymerase II holoenzymes *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 4587–4590 (1995).
19. Nonet, M., Scafe, C., Sexton, J. & Young, R. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol. Cell. Biol.* **7**, 1602–1611 (1987).
20. Cismowski, M., Laff, G., Solomon, M. & Reed, S. KIN28 encodes a C-terminal domain kinase that controls mRNA transcription in *Saccharomyces cerevisiae* but lacks cyclin-dependent kinase-activating kinase (CAK) activity. *Mol. Cell. Biol.* **15**, 2983–2992 (1995).
21. Hengartner, C. J. *et al.* Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol. Cell* **2**, 43–53 (1998).
22. Orphanides, G., Lagrange, T. & Reinberg, D. The general initiation factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
23. Koleske, A. J. & Young, R. A. An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466–469 (1994).
24. Kim, Y.-J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
25. Stargell, L. A. & Struhl, K. A new class of activation-defective TATA-binding protein mutants: Evidence for two steps of transcriptional activation *in vivo*. *Mol. Cell. Biol.* **16**, 4456–4464 (1996).
26. Akhtar, A., Faye, G. & Bentley, D. L. Distinct activated and non-activated RNA polymerase II complexes in yeast. *EMBO J.* **15**, 4654–4664 (1996).
27. Gonzalez-Couto, E., Klages, N. & Strubin, M. Synergistic and promoter-selective activation of transcription by recruitment of TFIID and TFIIB. *Proc. Natl Acad. Sci. USA* **94**, 8036–8041 (1997).
28. Cormack, B. P., Strubin, M., Ponticelli, A. S. & Struhl, K. Functional differences between yeast and human TFIID are localized to the highly conserved region. *Cell* **65**, 341–348 (1991).

Acknowledgements. We thank M. Benson, J. Geisberg, M. Harrison and S. Reed for DNAs and yeast strains; L. Stargell for TBP antibodies; M. Green for communicating unpublished results; and O. Aparicio, S. Bell and J. Lis for discussion. This work was supported by postdoctoral fellowships to L.K. from the Association pour la Recherche sur le Cancer and the Human Frontier Science Program, and by research grants to K.S. from the NIH.

Correspondence and requests for materials should be addressed to K.S. (e-mail: kevin@hms.harvard.edu).

Microscopic advances

Nature's annual New on the Market section devoted to matters microscopic and to image analysis includes knives and microtomes, and 'prescription' eyepieces to allow wearers of spectacles to use optical microscopes in comfort.

Finesse microtomes

From Shandon Lipshaw www.shandon.com

Ergonomically designed microtomes

Finesse microtomes are available with either manual or motorized cutting and are suitable for paraffin- or resin-embedded specimens. Both models feature electronic specimen advance, and allow the user to select specimen to stop compression of the block on the return stroke. Finesse ME (Motorized Electronic) features adjustable cutting speed. For delicate work such as sectioning resins, a stereo microscope or illuminated magnifier may be mounted on the Finesse.

Reader Service No. 100



Cut price: Shandon's Finesse microtome.

The Histo Knife

From Diatome www.emsdiasum.com

Sectioning between 0.2 and 10 µm

The Diatome HistoKnife produces thin sections for light and electron microscopy without the sound of breaking glass. The knife comes in 4, 5, 6, 8 and 10 mm diamond widths, using only natural diamonds. The knife may be used with hard and soft materials and works with all ultramicrotomes as well as regular microtomes with retractable knife holders in the return phase.

Reader Service No. 101

UltraVIEW

From EG&G Wallac www.wallac.com

A new microlens dual-disk for high confocal optical efficiency

The UltraVIEW confocal laser scanning microscope is aimed at the market for increased nonfocal optical resolution and fast imaging in the field of fluorescence cell biology — particularly green fluorescent protein (GFP) imaging. UltraVIEW is suitable for both fast and high resolution confocal microscopy. It allows up to 100 frames per second acquisition which can be used for true physiological applications in living cell research requiring high frame rates.

Reader Service No. 102

IdentiCheck AutoIMAGE

From Perkin-Elmer www.perkin-elmer.com

Fourier transfer-IR microscopy combined with the simplicity of NIR

The IdentiCheck AutoIMAGE Near Infrared microscopy system is built to give high-quality NIR reflectance spectra without time-consuming sample preparation. It can also readily measure NIR transmission spectra

through individual crystals with a spatial resolution below 10 µm. Designed as a complementary technique for both NIR spectroscopists and mid-IR spectroscopists, the IdentiCheck AutoIMAGE NIR microscope features complete system automation.

Reader Service No. 103

TruView IR Microscope

From Nicolet Instruments www.nicolet.com

Infinity-corrected optics for sharp images

The TruView Continuum IR microscope offers options from the basic microscope to a fully automated system with chemical mapping. The Continuum uses infinity-corrected optics with differential interference contrast to enhance contrast in samples such as biological cells and surface defects. Masking is achieved pre- and post-sample by a single motorized aperture.

Reader Service No. 104

AxioVision Version 2.0

From Karl Zeiss www.zeiss.de

Quantitative analysis and more functionality for digital microscopy

All motorized microscopes from Zeiss can be controlled by Axio Vision software. It provides biologists, physicians and also users in the fields of materials analysis and quality control with new options for digital image acquisition and processing, digital image and text archiving, and for the convenient compilation of reports. Two different operating modes are available: the user mode and the export mode. This enables the user to opt for simple operation for routine applications or high flexibility for experimental applications.

Reader Service No. 105

VectaMount

From Vector Laboratories www.vectorlabs.com

Toluene-free permanent mounting medium

VectaMount mounting medium is a toluene and xylene-free, non-flammable, non-hazardous, and odourless formula for permanently preserving tissue sections and cell preparations. This provides optimal mounting for a variety of histological, immunohistochemical, and *in situ* hybridization applications, retaining the colour and intensity of commonly used peroxidase and alkaline phosphatase substrates over time.

Reader Service No. 106

Micro-View Eyepieces

From MICRO 2000

Prescription eyepieces for your microscope

Now you can equip your microscope with custom made MICRO-VIEW eyepieces, made from your own ophthalmic prescription. Send MICRO 2000 your prescription and the size of your eyepieces. The eyepieces install in seconds, and since there is no permanent alteration, they can be removed as easily. Optically correct, they can reduce eyestrain, headaches and fatigue. In the States, call 800-850-2316, access code 00.

Reader Service No. 107

EM KMR 2 glass knifemaker

From Leica Microsystems Inc.

www.leica-microsystems.com

For plastic sectioning in light microscopy

The EMKMR2 uses the balanced break method to produce glass knives for in ultramicrotome and for plastic sectioning in light microscopy. The breaking mechanism is located just above the glass so a slow break can be achieved. The balanced break method gives a knife angle of 45°, necessary for producing ultra-thin cryosections.

Reader Service No. 108



Balancing act: Leica's EM KMR 2.

Photosensor Modules

From Hamamatsu www.hamamatsu.com

High sensitivity, wide dynamic range and fast time response for photon counting

Hamamatsu's HS773 and H5783 Ultra Compact Photosensor Modules combine an OPTO-8 PMT and power supply in a metal package. The modules operate from low voltage (+12V d.c.) and draw little current (180 mW). They are small enough to fit into spaces too tight for conventional PMTs, and even some semiconductor modules. The modules deliver high sensitivity, wide dynamic range, and fast time response and can be selected for photon counting. The H5773 mounts directly on a circuit board. The H5783 mounts on any surface and delivers its signal through a dedicated cable. Their compact size and low voltage operation combined with photon counting capabilities made these modules suitable for viewing sample plates in luminescence applications.

Reader Service No. 109

SZX12 Stereo Zoom Microscope

From Olympus www.olympus.com

For greater zoom and magnification ranges

The SZX12 Stereo Zoom Microscope features a zoom ratio of 12.8:1 ($0.7\text{--}9.0\times$), allowing the user to view complex structures in detail without changing the lenses. Using the dual nosepiece, a total stepless magnification range of $3.5\text{--}144\times$ is achievable, or $2.1\text{--}675\times$ with all objectives and eyepieces. A maximum numerical aperture of 0.13 is achieved with the Plan Apo $1\times$ objective. Olympus has developed an apochromatic auxiliary $20\times$ objective for the SZX12. With this attachment, stereo imaging is possible at magnifications up to $225\times$ with $10\times$ eyepieces, and $675\times$ with $30\times$ eyepieces. The user can achieve high contrast images with a resolution of 800 lines per mm. The system offers a wide range of parfocal objectives can



Zooming in: the SZX12 majors on flexibility.

be used in combination with brightfield, darkfield, oblique and fluorescence observation modes.

Reader Service No. 110

Montage Explorer

From Syncoptics www.syncoptics.co.uk

Digital imaging for optical microscopy

When the microscope stage is moved through the depth of field and around the sample under examination, Montage Explorer combines images from a microscope, building up high resolution pictures over a large field of view. In neurological tree studies, for example, axons can be traced throughout their length even in thick histological sections. And for entomologists, the ability to build a database of fully focused images of whole insects could open up new avenues in taxonomic research. Montage

Explorer allows materials scientists to map fractures over curved surfaces in any direction. The system is compatible with any optical microscope, without modifications. It includes an ink jet printer capable of producing photographic quality colour prints at up to A2 size.

Reader Service No. 111

Microscope Incubator

From Solent Scientific www.solentsci.co.uk

An incubator for use with the Olympus IX range of research inverted microscopes

Recent advances in image analysis software have encouraged more time-lapsed cell death and cell motility studies to be conducted. This acrylic chamber maintains the living cells at their optimum temperature, usually 37°C . Both transmitted-light and fluorescence experiments are supported. Other functions of the microscope; 35-mm camera, multi-viewing and micro-manipulation remain unaffected by the presence of the incubation chamber. Motorized stages enable several experiments to be studied simultaneously. An optional accessory maintains a CO_2 -enriched atmosphere which extends the life of the growth medium during long-term experiments.

Reader Service No. 112

OEM Filters

From Glen Spectra www.isa-gs.co.uk

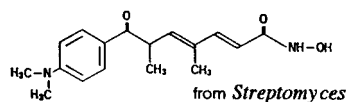
High-performance optical filters and coatings

Omega Optical produce high-performance optical filters and coatings for large-volume OEM applications. A wide range of components is available for building a "proof of concept" filter with a short delivery time. Following that, we can custom-design and produce filters for a variety of OEM requirements of spectral light measurement and control.

Reader Service No. 113

ADVERTISEMENTS

Trichostatin A



Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call
Wako BioProducts: (800) 992-9256; (804) 271-7677.

READER ENQUIRY NO. 98

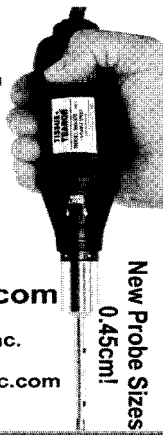
Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
 Ph/Fax 918-336-3363
 e-mail info@biospec.com



READER ENQUIRY NO. 54

Material, failure and quality analyses

atomic force microscopy (AFM), electron microscopy (STEM/SEM/SEM-EDS/SEM-CL) - fair prices
Laboratory Dr. Weischer
 Contract research - Seminars
 Schmidtbonnstr. 8,
 D - 53115 BONN, Germany,
 Tel. +FAX:0049(0)228-2125 44

READER ENQUIRY NO. 19

PriorLab microscope

From Prior Scientific

New laboratory microscope

The PriorLab laboratory microscope is available with a choice of PF achromatic or planachromatic objectives. Objectives are mounted in the rear-facing quadruple-position nosepiece, giving clear access to the specimen. A centreable, focusing condenser is provided along with a field lens assembly and diaphragm for Koehler-type illumination. A choice of viewing heads is available, including a trinocular version for photomicrography using 35-mm or CCTV cameras. The binocular and trinocular heads provide dual-dioptre adjustment to ensure parfocality at any interpupillary separation.

Reader Service No. 114**SMZ-800 Zoom Stereo Microscope**

From Nikon

www.nikonusa.com*A zoom stereo microscope promising comfort, optics and expandability*

The SMZ-800 Stereo Microscope series offers a new design, claimed to reduce fatigue. It's built for easy upgrading, with accessories that allow users to custom-configure their microscopes. Several accessories help bring the eyepieces to the user, instead of requiring the user to change posture. Of the three binocular tubes available, two are fixed-inclination tubes, a standard and a low position both at 20° inclination. The third ERGO-style binocular tube has low and high positions and has 30° of tilt. The focusable ERGO 1× objective lens allows the microscope to be placed in a comfortable position while magnification and working distance remain constant. The 6.3× zoom has click stops at 1× intervals from 2× to 5×, so operators can keep viewing the specimen while zooming. Additional features include fibre optic, halogen and ring illuminators, photomicrography and CCTV capability.

Reader Service No. 115**BAS-18000 Bio-Imaging Analyzer**

From Fujifilm

www.fujimed.com*A compact bio-imaging analyser*

The latest in the series of Bio-Imaging Analyzer Systems (BAS), this takes up less space, yet is claimed to give the same level of quantitative accuracy, resolution and linearity of data as bigger machines. The BAS-1800 was developed to combine accuracy and finely resolvable scanning with the sensitivity of Fujifilm Imaging Plates. These plates are based on a photostimulable phosphor layer (BaFX:Eu) and a flexible substrate more than 100 times more sensitive than the X-ray film it replaces. IP's reduce exposure time while increasing accuracy.

Reader Service No. 116

Prior engagement: open access to the specimen.

Variable pressure SEM

From Hitachi Scientific

www.inpress.co.uk*Microscopy of microorganisms*

Adding a cooling stage to a variable pressure SEM such as the Hitachi S-3000N increases the amount of time over which hydrated, uncoated samples can be examined. Use of a YAG back-scattered detector improves signal to noise and resolution at low accelerating voltages for these difficult samples. The performance of the combination is illustrated in Hitachi's technical data sheet No. 80, entitled "Improvement of image quality and microscopy of microorganisms".

Reader Service No. 117**Expanded Desktop**

From LEO

www.leo-em.co.uk*Two screens are better than one*

Expanded Desktop adds a second monitor to the SEM. It allows both monitors to act as a single Windows desktop, controlled from a single keyboard and mouse. The single Windows desktop allows applications to be located on either of the monitors separately, or even across both, all running and sharing data on a single PC. The option has been demonstrated running Microsoft Word on the second monitor for automatic report generation, and provides the platform for a fully integrated EDAX Phoenix EDS system, with the SEM running on one monitor and the analysis on the other.

Reader Service No. 118**Scanning Probe Microscopy**

From LOT Oriel

www.lot-oriel.com/uk*Applications for scanning probe microscopy*

Areas covered in a new series of Application Notes include drug delivery and carrier applications, and protein carriers such as liposomes studied at the submicrometre

level in buffers under varying environmental and temperature conditions. In the field of structural and molecular biology, a new application note covers the use of MAC Mode technology for imaging of cells such as DNA, proteins and chromatin at high, nanometre, resolution.

Reader Service No. 119**Technical Note 7**

From Bio-Rad

www.microscopy.bio-rad.com*Imaging deeper into living tissue using a short pulse laser for multi-photon fluorescence microscopy*

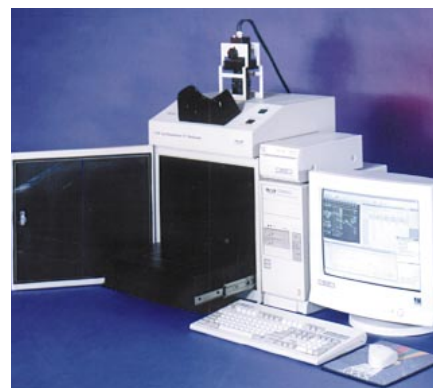
This technical note describes the advantages of using a short pulse laser for multiphoton fluorescence microscopy - a valuable technique for biomedical research. The method is proving useful for imaging deeper into biological tissue than any other available technique. By using femtosecond pulses to minimise the average power to a sample, damage to living cells is greatly reduced. Multi-photon fluorescence microscopy enables fluorescence excitation to be highly localised to the focal region of the laser. With no out of focus excitation; photobleaching and toxicity is minimal. In addition, there is no need for confocal collection optics, as it is possible to collect and image the fluorescence emitted, even if it is scattered as it escapes from the sample.

Reader Service No. 120**GDS-800 Chemi System**

From UVP Inc.

www.uvp.com*For chemiluminescence and low light applications*

This digital system from UVP is a stand-alone image acquisition and analysis system with a powerful PCI board which triggers a high-resolution cooled CCD camera. The system's Epi-Chemi Darkroom features a light-tight environment suitable for low light applications including chemiluminescence. Over-head UV (254 nm and 365 nm) is built-in along with white light. Labworks Software is also included with the GDS-8000 ChemiSystem.

Reader Service No. 121

Light program: the GDS-800 and ancillaries.

Openlab 2

From Improvision www.mcbainstruments.com
Software for controlling monochromators

This new software module in the Openlab 2 range is designed to take full advantage of the use of a monochromator. The user can scan the complete wavelength spectrum and preset up to four excitation wavelengths, which can be selected to match the precise peak excitation of each fluorescent probe, a particular advantage with GFP. Control of wavelength switching is provided by the Openlab Automator module. The interval between wavelengths is 3 ms, so that applications such as fast calcium imaging can be carried out.

Reader Service No. 122

MetaMorph 3.6

From Universal Imaging Corporation
www.image1.com

Multi-probe fluorescence added to image analysis software

The latest, 3.6 version of this software for image acquisition, analysis and processing, makes it easier to conduct multi-probe experiments using immunofluorescence, fluorescence *in situ* hybridization (FISH), and fluorescent proteins. A new feature allows the user to collect up to 6 wavelengths of images sequentially by controlling the camera and peripheral devices such as filter wheels and shutters. This function also provides auto exposure and correction of chromatic displacement.

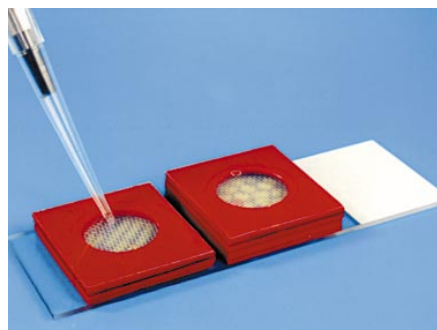
Reader Service No. 123

Whole-mount hybridization

From Grace Bio-Labs www.gracebio.com
On the slide...

This modular platform is made up of five interchangeable 'press-to-seal' silicone-based enclosures to secure thick and free-floating specimens in a small reagent volume. The system is designed to facilitate each step in the hybridization process, from permeabilizing, fixing and washing to staining and imaging, directly on a microscope slide without handling the specimen. Up to three enclosures can be built on one microscope slide.

Reader Service No. 124



Grace note: whole-mount hybridization.



SPOT colour: taking a cool look.

SPOT-2 Camera

From Diagnostic Instruments
www.diaginc.com

Microscope imaging camera with a dual position colour filter

The SPOT-2 camera is a cooled CCD colour digital camera designed specifically for microscopes. It can slide the colour and IR filter out of the light path to allow acquisition of true monochrome images. With the filters removed, higher wavelengths of light reach the CCD. This is essential for capturing light emission from Cy5 and Cy7 fluorescence dyes. The SPOT-2 camera includes a PCI card, power supply, cables, and a fully integrated software interface for acquisition, archival, processing, printing and annotation. The camera is compatible with Windows 95/98, Windows NT, and Mac operating systems.

Reader Service No. 125

Fluorescence Filter Guide

From Omega Optical www.omegafilters.com
A selection of stock filters and sets

Omega's 16-page "Fluorescence Filter Selection Guide for Microscopy" reflects current directions in biological and genetic research with the VIVID sets for GFP and the CyDyes and application-specific sets for GFP-FRET, FISH, and multiphoton work. For mainstream microscopy, the guide offers an assortment of neutral density filters, polarizing filters, and beamsplitters as well as special filters for IR Nomarski/DIC. Technical support sections include directions for choosing the right filter set and optimizing lamps and detectors to an expanded Fluorophore reference table of over 140 dyes.

Reader Service No. 126

Big Gel Imager 6240

From Ultra Lum www.ultralum.com
Modular gel documentation and image analysis system

The Ultra-Lum BGI-6240 I is a modular gel documentation/image analysis system built to handle gels in sizes up to 20 × 40 cm. All the cabinets in the 6000 Series feature a light-tight drop-down front door with a compression handle and safety interlock switch. The new interlock system protects against incidental UV exposure by turning off all UV light sources and allowing the white work light to stay on. BGI-6240 is ideal for DNA gels, protein gels, colony/plate counting, Dot/Slot blots, chemiluminescence and other types of gels.

Reader Service No. 127

Glass Based Dishes

From Bibby Sterlin www.bibby-sterlin.com

Glass-based dishes which allow improved fluorescent examination of cultivated cells

Incorporating either a glass or quartz base, these new dishes are recommended for observing live or dead cells with a nonfocal laser microscope. The materials used give high transmittance and low fluorescence, to make sensitive fluorescence measurements possible at high magnifications. Comprising an easy to handle 35 mm plate with either a 12 mm or 27 mm base attached with non-toxic silicon adhesive, the dishes combine a culture environment with the optical surface required for microscopy.

Reader Service No. 128

STFL2

From Micro Video Instruments
Attachment for fluorescence applications in stereo microscopy

The STFL2 epi-fluorescence attachment fits onto Nikon's SMZ-U stereo microscope to make the latest fluorescence labelling techniques available for the longer depth of field, lower magnifications and longer working distances that stereomicroscopy offers. The STFL2 attachment incorporates a mercury arc illuminator which provides intense excitation light. This passes through a heat absorbing filter before being reflected onto the specimen plane. Emitted fluorescence is gathered by the objective lens and filtered to screen off any residual excitation light. A filter changer holds up to three filter sets allowing the user to view three fluorescent labels without having to remove filters from the microscope.

Reader Service No. 129

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.